1998

Molecular systematics in Gossypium and its relatives

Tosak Seelanan
Iowa State University

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Molecular systematics in *Gossypium* and its relatives

by

Tosak Seelanan

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Botany (Systematics and Evolution)
Major Professors: Jonathan F. Wendel and Lynn G. Clark

Iowa State University
Ames, Iowa
1998

Graduate College
Iowa State University

This is to certify that the Doctoral dissertation of

Tosak Seelanan

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Co-major Professor

Signature was redacted for privacy.

Co-major Professor

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate College

TABLE OF CONTENTS

CHAPTER 1. GENERAL INTRODUCTION

The original objective of this research was to generate a phylogeny of *Gossypium* based on markers from the biparentally-inherited nuclear genome, and to compare this phylogeny to that based on the maternally-inherited chloroplast genome (Wendel and Albert 1992). As work progressed, a new question arose as to the monophyly of *Gossypium*. Therefore, more samples from species other than *Gossypium* had to be included. As a result, representatives from all but one genus in the tribe Gossypieae, to which *Gossypium* belongs, were included in the study. Although the taxonomic framework of the tribe Gossypieae had been laid out by previous authors (Alefeld 1861; Dumont 1887; Reeve 1936; Fryxell 1968, 1979), no formal phylogenetic inferences were made regarding relationships among members of this tribe (but see La Duke and Doebley 1995). Accordingly, phylogenetic investigation of the tribe Gossypieae as well as of *Gossypium* was conducted using molecular data from both chloroplast and nuclear genomes. Since the study employed multiple data sets, guidelines for data analyses were evaluated (e.g., Mason-Gamer and Kellogg 1996) and a new protocol was suggested (Seelanan et al. 1997).

The other aspect of the work presented in this dissertation is the study of the recently described *Gossypium* sect. *Grandicalyx* (Fryxell et al. 1992), from NW Australia. Recent expeditions to that region have provided a wealth of plant materials for studying this unique group of plants (Fryxell et al. 1992). The study described in this dissertation included broader sampling and employed more data than in previous studies (Wendel and Albert 1992; Seelanan et al. 1997).

To sum up, the primary objectives of this dissertation were:

1. To test the monophyly of the tribe Gossypieae, and to elucidate the phylogenetic relationships among genera within the tribe;
2. To investigate the monophyly of and relationships among species within the genus *Gossypium*, and to compare the results to a previously suggested phylogeny of the genus;
3. To study the phylogeny and evolution of Australian *Gossypium* sect. *Grandicalyx*.

Dissertation Organization

The main body of the dissertation is organized into four chapters: one chapter describing the major objectives and providing the relevant literature on molecular systematics; two chapters consisting of the original research results on issues of phylogenetic reconstruction and evolution of tribe Gossypieae and genus *Gossypium*; and a final chapter summarizing the results of this dissertation. All literature cited in Chapter 1 and 4 are listed after Chapter 4.

Chapter 2, entitled "Consensus and congruence in the cotton tribe (Malvaceae)," was published in the journal *Systematic Botany* (Seelanan et al. 1997). This chapter provides a practical solution to the long-standing problem of how multiple data sets from the same set of taxa can be used to infer organismal phylogenies. The approach presented was applied to a phylogenetic study of the cotton tribe [using *ndhF* and ITS data sets] and to *Gossypium* [using cpDNA restriction site variation from Wendel and Albert (1992) and ITS data sets]. Results from the phylogenetic study of the Gossypieae provide a new understanding of the relationships among its genera, its phytogeography, and the origin of the tribe. Results from the phylogenetic study of *Gossypium* reinforced our knowledge of the phylogeny of *Gossypium* as previously described (Wendel 1989; Wendel et al. 1991; Wendel and Albert 1992).

Chapter 3, entitled "Molecular systematics of Australian *Gossypium* L. sect. *Grandicalyx* (Fryxell) Fryxell (Malvaceae)," has been submitted to the journal *Systematic Botany*. This chapter provides insight to the phylogeny of *Gossypium* sect. *Grandicalyx* based on multiple molecular data sets. This section recently was taxonomically revised (Fryxell et al. 1992; Stewart et al. 1997) to include a total of 12 species from the Kimberley region of northwestern Australia. Three molecular data sets were employed in this research, namely, DNA sequences of the *rpl16* intron, nuclear 18S-26S ITS, and the nuclear alcohol dehydrogenase "locus D" gene. The results indicated that speciation of members in the section was relatively recent, so recent in fact that there was little molecular differentiation among this diverse assemblage of species. The results also provided a framework for interpreting the phytogeography, the origin of, and the morphological adaptation of *Gossypium* species in sect. *Grandicalyx*.

Chapter 4, entitled "General Conclusions," provides a summary of the work presented in this dissertation. In addition, other aspects for future study are suggested to further our knowledge of phylogeny in the Gossypieae and *Gossypium*.

Literature Review

Plant systematists have employed a variety of techniques to generate quality data for inferring organismal phylogeny. These include, for instance, morphological, anatomical, cytological, and molecular data. In particular, molecular data have been used intensively in the past and present decades, owing to the development of the Polymerase Chain Reaction (PCR) for DNA amplification and sequencing. This mini-review will present different types of molecular data that are used in molecular systematics, and will introduce the groups of plants to which some types of molecular data have been applied for phylogenetic purposes.

A still widely used source of molecular characters in plant systematics is chloroplast DNA (cpDNA) restriction fragment length polymorphisms (RFLP), or colloquially, cpDNA restriction-site variants. This type of data has been used to address questions at different taxonomic levels including the familial (e.g., Conant 1995; La Duke and Doebley 1995; Kohn et al. 1996), tribal and subtribal (e.g., Knox and Palmer 1995a; Yakawa and Uehara 1996), generic (e.g., Hoot 1995a; Knox and Palmer 1995b; Gillies and Abbott 1996; Ohnishi and Matsuoka 1996; Harrision et al. 1997; Olmstead and Palmer 1997; Spooner and Castillo 1997), and species levels (e.g., Mason-Gamer et al. 1995; Bain and Jansen 1996; Levy et al. 1996; Sewell et al. 1996). There are a few studies that employ PCR technology to generate portion(s) of cpDNA genes/introns/spacers, and use PCR-amplified products in RFLP surveys to address phylogenetic questions at different taxonomic levels (e.g., at the family level in Tsumura et al. 1996; at the tribal level in Wolf et al. 1997; at the species level in Mousadik and Petit 1996).

Another type of molecular data is DNA sequences from the chloroplast genome. Although cpDNA on the whole has a slow rate of nucleotide substitution compared to other genomes in plants, different parts of the genome have different rates of evolution (Palmer 1991), and can be used to

address phylogenetic questions at different levels. Many chloroplast genes have been used in phylogenetic studies including, for example, *rbcL* (e.g. Chase et al. 1993; Rodman et al. 1993; Rodman et al. 1996; Duvall and Morton 1996; Soltis et al. 1996; Wagstaff and Olmstead 1997), *ndhF* (e.g. Clark et al. 1995; Neyland and Urbatsch 1996; Bohs and Olmstead 1997; Terry and Brown 1997; Terry et al. 1997), and *matK* (e.g., Johnson et al. 1996; Liang and Hilu 1996). Introns and spacers are also used. They often provide more information than gene sequences, for example, the *rpoC1* intron (Downie et al. 1996), *rpl2* intron (Bailey and Doyle 1997), *trnL* (UAA) intron (Gielly and Taberlet 1996), *atpB-rbcL* spacer (Savolainen et al. 1997), and *rpl16* intron (Kelchner and Clark 1997). In many studies, more than one gene/intron/spacer have been used (e.g., Hoot 1995b; Olmstead and Reeve 1995; Soltis et al. 1996; Plunkett et al. 1997).

In addition to cpDNA restriction-site variation and sequences, structural rearrangements, due to loss of gene/intron/inverted repeat or inversions, can be useful for phylogenetic and evolutionary studies at the family level (e.g., Doyle et al. 1995; Doyle et al. 1996; Katayama and Ogihara 1996; Wallace and Cota 1996). This type of data, although potentially informative, is limited in terms of the numbers of characters that it provides, and it may be homoplastic. A critical review of cpDNA systematics with respect to the advantages and disadvantages of each technique has recently been published (Olmstead and Palmer 1994), and therefore it will not be repeated here.

Mitochondrial DNA (mtDNA), compared to cpDNA, is a much less versatile molecule due to a high rate of rearrangement and a low rate of point mutation (Palmer 1992). Nonetheless, a few studies have used mtDNA to infer phylogenetic relationships among and evolution in taxa studied (e.g., Lou and Boutry 1995; Guo et al. 1996; Mori et al. 1997).

Many types of molecular data have been generated from the nuclear genome, but the most widely used are RFLP's and sequences of the ribosomal DNA (rDNA) gene [including 18S, 26S, 5S rDNA genes, and internal transcribed spacer (ITS) regions between 18S and 5.8S and between 5.8S and 26S]. RFLP's of the rDNA gene have been used to address the origin of polyploid species (e.g., Doyle and Brown 1989), intra- and interspecific variation (e.g., Bobola et al. 1992), and phylogenetic

reconstruction (e.g., Sytsma and Schaal 1985; Bellarosa et al. 1990; Reddy et al. 1990). DNA sequences of rDNA genes are usually used largely at higher taxonomic ranks (above the family level) (Kron 1996; Potter et al. 1997; Soltis and Soltis 1997; Kuzoff et al. 1998). ITS sequences are used more widely to infer phylogenetic relationships among species within a genus (e.g., Baldwin 1993; Wojciechowski et al. 1993; Sun et al. 1994; Yuan et al. 1996; Holst-Jensen et al. 1997) and within a tribe or higher taxonomic ranks (e.g., Savard et al. 1993; Suh et al. 1993; Bogler and Simpson 1996), to document reticulate evolution (e.g., Wendel et al. 1995; Sang et al. 1995), and to infer morphological evolution from ITS-based phylogeny (e.g., Desfeux et al. 1996). It should be noted that when the rDNA gene and spacer are employed, caution must be taken because molecular evolution governing this multigene family can give a gene tree that differs from the organismal tree (Wendel et al. 1995). Accessing orthologous/paralogous relationships of rDNA genes and spacers may be problematic, as rDNA loci can move from one location on a chromosome to another or even to a different chromosome in an evolutionarily brief time span (Dubcovsky and Dvorak 1995).

Only recently, systematists have started using genes from other small multigene families, for example, those from nuclear alcohol dehydrogenase (Small et al. 1998), and nuclear phosphoglucose isomerase families (Gottieb and Ford 1996). Only well-characterized nuclear genes should be used, and caution must be exercised in order to ensure that phylogenetic reconstruction is based on orthologous copies, not paralogous ones.

In this dissertation, a variety of molecular data types were used, including cpDNA restriction-site variation, chloroplast *ndhF* gene and *rpl16* intron sequences, and nuclear rDNA ITS and alcohol dehydrogenase (*Adh*) gene sequences. The chloroplast *ndhF* gene and the nuclear ITS of rDNA were used to reconstruct the phylogeny of the tribe Gossypieae of the family Malvaceae. The cpDNA restriction-site variation and nuclear ITS of rDNA gene were used in a phylogenetic investigation of *Gossypium*. The *rpl16* intron, and nuclear ITS of rDNA and alcohol dehydrogenase (*Adh*) genes were used in cladistic analyses of Australian *Gossypium* sect. *Grandicalyx*.

Currently, the Malvaceae, or mallow family, consists of approximately 75 genera (Zomlefer 1994), including many economically important crops

such as cotton, okra, and ornamental plants (e.g., Portia plant, hollyhock, mallow). The family is divided into 5 tribes: Malveae, Malvavisceae, Hibisceae, Gossypieae, and Decaschistieae (Fryxell 1988). These tribes can be distinguished by a combination of many characters including fruit type, number of mericarps per fruit, style connation, stigmas:mericarp ratio, apex of staminal column, and presence of gossypol glands [lysigenous cavities] (Fryxell 1988). To date, there are a few phylogenetic works at the familial level (e.g., La Duke and Doebley 1995), but none directed at resolving tribal relationships.

Among the five tribes described in the Malvaceae, the Gossypieae, the cotton tribe, is the most distinctive, having distinct lysigenous cavities capable of synthesizing terpenoid aldehydes throughout the plant, and unique embryo morphology, traits recognized by Alefeld (1861) as characteristic of Gossypideae [Gossypieae]. Many authors (e.g., Bentham and Hooker 1862-1867; Schumann and Gürke 1891-1892; Edlin 1935; Kearney 1951; Corner 1976) had overlooked these features and preferred to include members of this tribe in the tribe Hibisceae. Fryxell (1968) resurrected the tribe Gossypieae on the basis of Alefeld's (1861) finding, with other supporting evidence from wood anatomy (Dumont 1887) and seed coat anatomy (Reeve 1936). In his critical review, Fryxell (1968) included eight genera within the tribe Gossypieae, viz., *Gossypium* L., *Gossypioides* Skovst. ex J. B. Hutch., *Kokia* Lewton, *Thespesia* Sol., *Hampea* Schltdl., *Cienfuegosia* Cav., *Lebronnecia* Fosberg, and *Cephalohibiscus* Ulbr. Several of these genera consist of few species and are confined to narrow geographic distributions, such as *Cephalohibiscus* from New Guinea, *Lebronnecia* from the Marquesas Islands (Fosberg and Sachet 1966), *Gossypioides* from southern Africa-Madagascar (Hutchinson 1947), and the Hawaiian endemic genus *Kokia* (Degener 1934). The Mexican endemic *Hampea* (Fryxell 1969a, 1988) has a relatively wider geographic distribution and includes 21 species. The other three genera are pantropical. Twenty-five and 17 species, respectively, are included in *Cienfuegosia* (Fryxell 1969b) and *Thespesia* (Fryxell 1968). The biggest and most diversified genus with the widest geographic distribution is *Gossypium* (Fryxell 1992; Stewart 1995; Stewart et al. 1997) with more than 50 species described.

To date, there have been no phylogenetic studies of relationships among genera within the tribe. A recent phylogenetic study of the Malvaceae (La Duke and Doebley 1995) included only three representatives from the tribe, and gave little indication of relationships among taxa. Clearly, a lack of phylogenetic study in this group needs to be addressed to further our understanding of the relationships among genera within this economically important tribe.

Among the eight genera in the tribe Gossypieae, *Gossypium* receives the most attention from many authors as four members of the genus are economically significant crops providing cotton to the world. These commercially important cottons include two diploid species, *Gossypium arboreum* L. and *G. herbaceum* L., and two tetraploid species *G. hirsutum* and *G. barbadense*. Many studies have been done on these plants, including investigations of taxonomy (e.g., Fryxell 1992; Fryxell et al. 1992; Stewart et al. 1997), cytogenetics (e.g. Endizzi et al. 1985; Price et al. 1990; Crane et al. 1993), domestication (e.g., Wendel et al. 1992; Brubaker and Wendel 1994), phylogenetics and evolution (e.g. Wendel 1989; Wendel et al. 1991; Wendel and Albert 1992; Wendel et al. 1995a; Seelanan et al. 1997; Small et al. 1998), and molecular evolution (e.g. Wendel et al. 1995b; Cronn et al. 1996). Despite the extensive studies on this genus, new discoveries have recently emerged (Fryxell et al. 1992; Stewart 1995), for example, the new group of *Gossypium* species that is found only in Kimberley region in northwestern of Australia, with the exception of one species that occurs 500 km away on the Cobourg Peninsula. These species possess a number of morphological characters that are found nowhere else in the genus, including their habit as upright or prostrate perennial herbs, the ability to generate new shoots from rootstocks after above-ground vegetative parts are burned by fires, and adaptation of reproductive parts to facilitate ant dispersal. Previous studies (Wendel and Albert 1992; Seelanan et al. 1997) have focused on generic level relationships or groups other than Australian *Gossypium* (but see Wendel et al. 1991). Here, this group is the focus of an intensive phylogenetic study.

CHAPTER 2. CONGRUENCE AND CONSENSUS IN THE COTTON TRIBE (MALVACEAE)

A paper published in the journal *Systematic Botany*[1]

Tosak Seelanan[2,3], Andrew Schnabel[4], and Jonathan F. Wendel[5]

ABSTRACT

We explored the evolutionary history of the Gossypieae and *Gossypium* using phylogenetic analysis of biparentally and maternally inherited characters. Separate and combined data sets were analyzed and incongruence between data sets was quantified and statistically evaluated. At the tribal level, phylogenetic analyses of nuclear ribosomal ITS sequences yielded trees that are highly congruent with those derived from the plastid gene *ndhF*, except for species that have a reticulate evolutionary history or for clades supported by few characters. Problematic taxa were then pruned from the data sets and the phylogeny was inferred from the combined data. Results indicate that (1) the Gossypieae is monophyletic, with one branch from the first split being represented by modern *Cienfuegosia*; (2) *Thespesia* is not monophyletic; and (3) *Gossypium* is monophyletic and sister to an unexpected clade consisting of the Hawaiian genus *Kokia* and the east African/Madagascan genus *Gossypioides*. Based on the magnitude of *ndhF* sequence divergence, we suggest that *Kokia* and *Gossypioides* diverged from each other in the Pliocene, subsequent to their apparent loss of a pair of chromosomes via chromosome fusion. Phylogenetic relationships among species and "genome groups" in *Gossypium* were assessed using cpDNA

---

[1] Reprinted with permission of the journal *Systematic Botany*, 1997, 22: 259-290.

[2] Graduate student, Department of Botany, Iowa State University

[3] Primary researcher and author

[4] Associate Professor, Department of Biological Sciences, Indiana University

[5] Professor, Department of Botany, Iowa State University

restriction site variation and ITS sequence data. Both data sets support the monophyly of each genome group, once taxa known or suspected to have reticulate histories are pruned from the trees. There was little congruence between these two data sets, however, with respect to relationships among genome groups. Statistical tests indicate that most incongruence is not significant and that it probably reflects insufficient information rather than a biological process that has differentially affected the data sets. We propose that the differing cpDNA- and ITS-based resolutions of genome groups in *Gossypium* reflect temporally closely spaced divergence events early in the diversification of the genus. This "short internode" phenomenon is suggested to be a common cause of phylogenetic incongruence.

INTRODUCTION

One of the issues that systematists face when multiple data sets exist for a particular phylogenetic problem is whether or not to combine these different sources of information into a single analysis (Miyamoto 1985; Kluge 1989; Barrett et al. 1991; Hillis 1987; Bull et al. 1993; de Queiroz et al. 1995; Miyamoto and Fitch 1995). This question stems in part from the recognition that different character sets can have different underlying evolutionary histories, which may lead to different reconstructions of organismal phylogenies. In some cases, these different histories may be clear or at least strongly suggested by the available information, as with the well-known phenomenon of cytoplasmic introgression in plants that leads to incongruent nuclear and cytoplasmic gene trees (Rieseberg and Soltis, 1991). Perhaps more common, though, are situations where the biological or analytical sources of incongruence are obscure. Irrespective of the origin of the incongruence, one possible remedy is to remove from the data sets the taxon or taxa displaying the incongruence and conduct a combined analysis on the remaining taxa. This process, however, necessitates that one first assess whether topological differences between competing phylogenies are real (Templeton 1983; Faith 1991; Farris et al. 1995; Mason-Gamer and Kellogg 1996; Kellogg et al. 1996) and hence potentially biologically significant, or whether they reflect insufficient character evidence, excessive homoplasy, or some other cause of spurious reconstruction. The question of whether or not different data sets should be

combined into a single global analysis is thus intertwined with the issue of interpreting topological differences.

In this paper we present an example of both aspects of this problem using different data sets for the malvaceous tribe Gossypieae and for the genus *Gossypium*. The Gossypieae was first distinguished from other Malvaceae on the basis of embryo morphology and by the presence of punctae or lysigenous cavities (colloquially called "gossypol glands") that are widely distributed throughout the plant body (Alefeld 1861). The taxonomic value of these characters has not been universally recognized, and members of the tribe have often been included with other capsular-fruited genera in the Hibisceae (e.g., Bentham and Hooker 1862-1867; Edlin 1935; Kearney 1951). The tribe was resurrected by Fryxell (1968), who through comparative studies of the Gossypieae and Hibisceae recognized the significance of embryo morphology, the presence of gossypol glands, and additional characters from wood anatomy (Dumont 1887) and especially seed coat anatomy (Reeves 1936). Members of the the two tribes differ, for example, in embryo conformation (simple folding in the Hibisceae vs. complex folding in Gossypieae) and number of cell layers in the outer integument (two vs. many). Both the close affinity of the two tribes and their distinctness are evident in Fryxell's (1968) tabulation and discussion of morphological characters, and by recent phylogenetic analysis of cpDNA restriction site variation (La Duke and Doebley 1995).

Eight genera presently are recognized in the Gossypieae (Fryxell 1968, 1979). The cpDNA restriction site phylogeny of La Duke and Doebley (1995) suggests that the tribe is monophyletic, although only one species was sampled from each of three genera. Four of the eight genera are small with restricted geographic distributions. *Lebronnecia* Fosberg is a monotypic island endemic from the Marquesas (Fosberg and Sachet 1966). *Cephalohibiscus* Ulbr., from New Guinea and Solomon Islands, also is monotypic (Fryxell 1968). Two species, from East Africa and Madagascar, are described in *Gossypioides* Skovst. ex J. B. Hutch. (Hutchinson 1947). The Hawaiian endemic *Kokia* Lewton includes four species, of which one is extinct (Lewton 1912; Degener 1934; Hutchinson 1947). In addition to these four small genera, the tribe includes four moderate sized genera with broader geographic ranges. *Hampea* Schltdl., the only genus in the tribe

with dioecious members (except perhaps *Cienfuegosia heteroclada* Sprague; Fryxell 1979) comprises 21 neotropical species (Fryxell 1969a, 1979). In the most recent revision of *Cienfuegosia* Cav., Fryxell (1969b) recognized 25 species, with an aggregate range that includes the neotropics and parts of Africa. Seventeen species are recognized in the pantropically distributed *Thespesia* Sol. (Fryxell 1979).

The largest and most widely distributed genus in the tribe is *Gossypium* L., which contains more than 50 species (Fryxell 1992), including four (*G. arboreum* L., *G. herbaceum* L., *G. barbadense* L., *G. hirsutum* L.) that attract attention by virtue of their economic importance. *Gossypium* is distinguished from related genera by a combination of characters (undivided style, coriacious capsule containing several seeds per locule, somatic chromosome number of 26, and three foliaceous [usually] involucellar bracts), each of which is found in related genera. Thus, although the genus has been studied from many perspectives, no unique synapomorphic character is known to define *Gossypium*, and its presumptive monophyly is untested. The sole phylogenetic hypothesis for the genus is based on characters from the maternally inherited (in *Gossypium*) plastid genome (Wendel and Albert 1992).

Although the Gossypieae has recently been revised (Fryxell 1979) and a skeletal phylogenetic hypothesis for the family has been erected (La Duke and Doebley 1995), relationships among genera have not been formally evaluated and the monophyly of the tribe remains untested. In the present study, we conducted phylogenetic analyses of both biparentally (nuclear) and maternally (plastome) inherited markers in an effort: (1) to test the monophyly of the Gossypieae and investigate generic-level relationships within the tribe, and (2) to test the monophyly of *Gossypium* and investigate species-level relationships within the genus. In the course of this work, we generated different data sets at both the tribal and generic levels, using parallel sampling of taxa. Phylogenetic analyses revealed broad areas of agreement in trees derived from the different data sets, in addition to regions of incongruence. We discuss the implications of the results that are congruent, and also consider the topological incongruence in an effort to diagnose its causes and to evaluate competing phylogenies.

MATERIALS AND METHODS

Plant Materials. The accessions selected for study are listed in Table 1. Accessions were derived from seeds of plants growing in their native habitat, from material derived from self-pollination of original field-collected plants, or as gifts from various collections. For most accessions, plants were grown from seed to verify identity, and voucher specimens, where available, were deposited in the Ada Hayden Herbarium (ISC) at Iowa State University. We included one or more representatives of seven of the eight genera in the Gossypieae, lacking only the rare monotypic taxon *Cephalohibiscus peekelii* Ulbr.. Sampling density varied widely among the other ingroup genera: *Lebronnecia kokioides*, the sole representative of this monotypic genus; one species from each of two sections (of 26 species in the genus) in *Cienfuegosia*; two of four species from *Kokia*; three species (of 17) from the two sections of *Thespesia*; one of 21 species of *Hampea*; one of two species from *Gossypioides*; and 52 accessions of 45 species from *Gossypium*, representing the majority of the diversity in the genus (several poorly known African-Arabian taxa were not available).

Outgroups were selected for inclusion based on the morphological evidence discussed above and the cpDNA restriction site analysis of La Duke and Doebley (1995). The latter analysis indicated that a clade consisting of the tribes Hibisceae and Malvavisceae is sister to the tribe Gossypieae. Two taxa, *Hibiscus costatus* from the Hibisceae, and *Anotea flavida* from the Malvavisceae, were therefore selected as outgroups for the Gossypieae.

DNA Extraction and Sequencing. DNA was extracted from freshly harvested or from silica-gel dried leaves using previously detailed methods (Paterson et al. 1993; Brubaker and Wendel 1994). In most cases, sequence data were obtained through direct sequencing of single-stranded templates derived from a two-stage polymerase chain reaction (PCR) procedure followed by ultrafiltration, as previously described (Clark et al. 1995; Wendel et al. 1995a). In some instances, however, it was necessary to generate sequence data from individual clones of the amplification products.

Primers for PCR amplification and DNA sequencing of *ndhF* are specified in Olmstead and Sweere (1994), although "primer 1" was redesigned to correspond to coordinates 114,183 - 114,164 of the tobacco

ndhF gene (Shinozaki et al. 1986) as follows: 5'-GAATATGCATGGATCATACC- 3'. Sequence data for ITS were generated as described in Wendel et al. (1995a), although only the internal primers ("ITS2" and "ITS3") located in the highly conserved 5.8S RNA gene were used for sequencing reactions in most cases. This allowed complete sequence data to be obtained from both spacers (ITS1 and ITS2) and a portion of the 5.8S gene in all taxa. When direct sequencing of PCR products failed to yield clean sequence, we used the pGEM®-T Vector (Promega) to clone PCR-amplified ITS fragments. Cloning followed the manufacturer's instructions, and individual clones served as templates for sequencing. Manual DNA sequencing was performed using Sequenase version 2.0 (Amersham), following the manufacturer's instructions, and electrophoresis in 1.2x TBE/ 5% polyacrylamide gels (Long Ranger, FMC). In some cases, we used either an ABI PRISM 377 DNA Sequencer or an ABI PRISM 310 Genetic Analyzer (Perkin-Elmer) for automated sequencing.

Data Analyses. The 19 *ndhF* sequences generated were readily aligned by eye, because there was relatively little length variation and the inferred amino acid sequences provided sufficiently confident assessments of homology. ITS sequences were aligned using the module PILEUP of GCG (Program Manual for the Wisconsin Package, Version 8, Genetics Computer Group, Madison) and the resulting alignments were manually adjusted as necessary.

PAUP version 3.1 (Swofford 1993) was used for most phylogenetic analyses. For the *ndhF* data set, BRANCH and BOUND searching was conducted using four different treatments of gaps: gapped regions omitted; gaps coded as missing; gapped regions omitted but added as presence/ absence characters; gaps coded as missing but added as presence/absence characters. In the ITS data set, gaps were coded as missing. Because of the large number of sequences in the ITS data set, HEURISTIC searches were necessary. To explore the possibility of multiple islands of trees (Maddison 1991), we used 1000 random addition sequences, with TBR branch swapping and MULPARS on. Ten trees from each replicate were retained for subsequent searching using TBR branch swapping until completion. In addition to the *ndhF* and ITS data sets, we included and re-analyzed a matrix of cpDNA restriction site variation from an earlier study (Wendel and

Albert 1992). This matrix consists of data for a single outgroup taxon (*Thespesia populnea*) and 39 *Gossypium* species, each with 202 binary restriction site characters.

Strength of support for each resolved clade was evaluated by generating decay indices (Donoghue et al. 1992), which express the number of step(s) that must be added to make each clade of interest disappear in a strict consensus of near-most-parsimonious trees. To obtain each decay index, we conducted searches for the shortest trees that failed to exhibit each monophyletic clade of interest (i.e., were constrained *not* to exhibit monophyly). The differences in tree length between the original and constrained searches represent the decay index. To facilitate these calculations, we used Autodecay version 3 (Eriksson, T. and N. Wikström, 1995. AutoDecay ver. 3.0 [Computer program distributed by the authors]. Botaniska Institutionen, Stockholm University. Stockholm.) in conjunction with PAUP.

Because comparisons of trees generated by the three different data sets showed incongruence, we performed additional analyses in which taxa known or suspected to have a reticulate evolutionary history were excluded. The justification for these added analyses was to exclude obvious sources of phylogenetic incongruence from subsequent comparisons among trees generated from *ndhF*, ITS, and cpDNA restriction site data sets. Taxa to be excluded were chosen based on previous phylogenetic work, as follows: (1) Several species have strikingly incongruent maternal and bi-parental phylogenetic placements, perhaps reflecting, in each case, plastome introgression or a bi-phyletic origin. These species are *G. bickii* (Wendel et al. 1991), *G. cunninghamii* (Wendel and Albert 1992), and the Colima populations of *G. aridum* (DeJoode 1992; Wendel and Albert 1992). (2) Five *Gossypium* species (*G. hirsutum*, *G. barbadense*, *G. tomentosum*, *G. darwinii*, and *G. mustelinum*) are allopolyploids containing Asiatic "A" and American "D" nuclear genomes in an A-genome-like cytoplasm (Endrizzi et al. 1985; Wendel 1989). Intergenomic concerted evolution has homogenized A-genome and D-genome sequences, thereby obscuring paralogy/orthology relationships (Wendel et al. 1995b). (3) Finally, we recently presented evidence documenting an unusual recombinant history for *G. gossypioides*, involving intergenomic hybridization between A-genome and D-genome taxa.

This process was invoked to account for a mosaic ITS sequence in *G. gossypioides* (Wendel et al. 1995a). In summary, nine species were excluded from some analyses.

Incongruence between data sets was evaluated by inspecting topologies for conflict and strength of support for individual branches, and was quantified using two indices, I_{MF}, the Mickevich-Farris index (Mickevich and Farris 1981), and I_M, hereafter the Miyamoto index (M. Miyamoto pers. comm. to Kluge 1989). Both indices may be interpreted as providing estimates of the proportion of characters that are in conflict between two data sets. Typically, the value of I_{MF} is roughly half that of I_M (Swofford 1991; Kim and Jansen 1994). Both indices were calculated, as described in Swofford (1991), for all relevant pairwise comparisons among the *ndhF*, ITS, and cpDNA data sets.

In addition to assessing overall conflict among data sets, specific topological alternatives were statistically tested for congruence using Wilcoxon Signed Rank (WSR) tests (Siegel 1956; Templeton 1983), which in the present context test whether one data set significantly conflicts with an alternative data set. As discussed by Mason-Gamer and Kellogg (1996), WSR tests have several attractive features relative to other techniques for statistically testing incongruence between data sets, such as the topology-dependent permutation tail probability test (Faith 1991) and the incongruence length difference test (Farris et al. 1995). Perhaps most important in this respect is the wide applicability of the WSR tests to topological levels ranging from entire trees to individual nodes. Thus, WSR tests allow statistical evaluation of incongruence on a local level. A potential limitation of these tests is that they take into account the number of *characters* exhibiting changes rather than differences in the *number of steps*. In principle, this may lead to situations where WSR tests are too conservative (failure to detect incongruence), for example, in cases where characters vary greatly in the number of steps between constraint trees and most-parsimonious trees. For instance, if five characters exhibit fewer steps and five require more steps, the WSR test statistic will be insignificant, regardless of how many *total* step differences are involved for the characters in question. In practice, this theoretical limitation is not problematic, as the

number of step differences for each character between shortest and constraint trees is typically low.

WSR tests were conducted using PAUP. First, separate searches for the most-parsimonious trees were executed for each data set and separate strict consensus trees were calculated. Second, for each node or nodes of interest, the topology from one data set was used as a constraint for the alternative data set and a new tree search was conducted. Third, the list of character changes in trees obtained in both analyses was examined, and differences in the number of steps required for individual characters were tabulated: some characters required fewer steps on the constraint tree while other characters required more steps. Finally, for each tabulation, WSR tests were implemented as described in Siegel (1956), and the associated probabilities were obtained from statistical tables (Gibbons 1985; Harter and Owen 1970).

In performing the WSR tests, we followed the recommendations of Mason-Gamer and Kellogg (1996) and Kellogg (pers. comm.) and used "conservative" constraint trees, i.e., those that were resolved only to the extent necessary to test the node of interest. For example, if one data set yields a phylogeny (OG,(((a,b),((c,d),e)),((f,g),((h,i),(j,k))))), and one wishes to evaluate whether (a,b) is sister to (c,d,e), the appropriate constraint is ((a,b),(c,d,e)). If, on the other hand, the node of interest is "a through e" monophyly, the appropriate constraint would impose no internal structure for this clade, and therefore would be (a,b,c,d,e).

RESULTS

Variation in ndhF. After eliminating regions corresponding to the amplification primers, which correspond to coordinates 114,183 - 114,164 and 112,179 - 112,154, respectively, of the tobacco gene (Shinozaki et al. 1986, as corrected by Olmstead et al. 1993), and two small regions of poor resolution (coordinates 114,129 - 114,124 and 114,150 - 114,148), sequences corresponding to 2058 nucleotides of the tobacco *ndhF* gene (114,162 - 112,180, less 112,686 - 112,684 and 112,647 - 112,645) were generated. The 19 sequences were easily aligned by eye with the introduction of five in-frame gaps, all either one or two codons in length. The homoplasious two-codon insertion "b" (Figs 1, 2) is associated with a

short direct repeat, as also reported in the Asteraceae (Kim and Jansen 1995) and Solanaceae (Olmstead and Sweere 1994). This suggests a possible mechanistic basis for the observed homoplasy (see also Kelchner and Wendel 1996). Accommodating the gaps in the Gossypieae data set lengthened the matrix by a total of 12 nucleotides, resulting in a final matrix of 19 rows and 2070 columns.

Within the Gossypieae, excluding polymorphisms introduced by gaps, 127 sites (6.1%) were variable, of which more than half (59.8%) were third codon positions (76 sites). An additional 22.1% and 18.1% of the variable positions were at first and second codon positions (28 and 23 sites, respectively). Approximately one-third of the 127 variable characters were potentially phylogenetically informative (46 sites). For the gene as a whole, 2.2% of the characters yielded potentially phylogenetically informative variation in the tribe, compared to 4.8% in the Solanaceae (Olmstead and Sweere 1994), 16.8% in the Asteraceae (Kim and Jansen 1995), 19.6% in the Acanthaceae (Scotland et al. 1995), 22.9% in the Poaceae (Clark et al. 1995), and 26.0% in the Scrophulariaceae (Olmstead and Reeves 1995). Variability in *ndhF* is approximately 1.4 times as high in the 3' half of the gene as in the 5' half of gene (74 vs. 53 sites), with a slightly greater bias (29 vs. 17) for sites that are potentially phylogenetically informative (Fig. 1). This distribution pattern of variable and potentially informative sites is similar to that observed in other studies, although the bias for increased variability in the 3' half of the gene is often greater. As shown in Fig. 1, the insertions/deletions (indels) observed in the Gossypieae also occur in the 3' half of the gene, consistent with tendencies observed in other families (Olmstead and Sweere 1994; Clark et al. 1995; Kim and Jansen 1995).

Averaged across all pairwise comparisons in the tribe, synonymous substitution (K_S) rates per 100 sites equal 2.7, whereas nonsynonymous substitution (K_A) rates are 0.8. Total substitution (K_O) rates average 1.23 among the 17 ingroup sequences. To place the magnitude of these estimates in a comparative context we computed substitution rates for *rbcL* using a subset of the taxa sampled for *ndhF*: *Thespesia populnea* (Genbank no. L01961); *Gossypium robinsonii* (L13186); *Gossypium hirsutum* (M77700); and *Gossypium anomalum* and *Gossypium davidsonii* (unpublished data). As shown in Table 2, substitution rates for all three classes of nucleotide

substitution are generally comparable for *ndhF* and *rbcL*. In comparisons involving *G. robinsonii* and *G. anomalum*, however, K_S is lower for *ndhF* (0.21) than for *rbcL* (1.5). The overall similarity of substitution rates for *ndhF* and *rbcL* contrasts with expectations based on higher substitution rates found in other studies (e.g., in the Asteraceae; Kim and Jansen 1995). Percent GC contents for the two genes are nearly identical in the two families (32.3% and 44.0% for *ndhF* and *rbcL*, respectively).

Phylogenetic analysis of ndhF sequences. For phylogenetic analyses, trees were rooted using sequences from *Hibiscus costatus* and *Anotea flavida*, both in combination and individually, so that ingroup stability could be assessed relative to outgroup choice. In all analyses, ingroup topology was not affected by outgroup selection, and the alternative outgroup was never nested within the ingroup. This latter result provides support, albeit weak due to the limited sampling, for a monophyletic tribe Gossypieae. For all analyses discussed here, both outgroup taxa were used to root trees.

Character-based phylogenetic analyses were conducted in four separate analyses that varied according to treatment of gaps. In the first analysis, where gapped positions were treated as missing data, 15 most-parsimonious trees were recovered (114 steps; CI = 0.83; RI = 0.90). In this analysis, the placement of *Cienfuegosia tripartita* was uncertain, and is shown in Fig. 3 (top) as variously resolving at positions denoted by **A**, **B** or **C**. In each of these trees, *Thespesia thespesioides* resolved as sister to *Lebronnecia kokioides* at the terminal denoted by the letter **i** (Fig. 3). In topology "**A-i**", *Cienfuegosia tripartita* represents the basal clade of the tribe, whereas in the other two topologies *Cienfuegosia* is sister to groups of three genera, either *Gossypium*, *Gossypioides* and *Kokia* ("**B-i**") or *Thespesia*, *Hampea*, and *Lebronnecia* ("**C-i**"). For each of these three differing placements of *Cienfuegosia*, five alternative arrangements were found within the genus *Gossypium* (G1-G5, inset of Fig. 3).

In the second analysis, gapped positions were excluded from the data matrix. Twelve most-parsimonious trees were recovered (113 steps; CI = 0.83; RI = 0.90), with topologies identical to topologies **A-i**, **B-i**, and **C-i**. For each of these topologies, four different arrangements were found in *Gossypium*, corresponding to G1, G2, G3, and G5 in the inset of Fig. 3.

Thus, the trees resulting from this analysis represent a subset of those recovered in the first analysis.

In the third analysis, in which gaps were added to the data matrix as presence/absence characters in addition to being coded as missing data, ten most-parsimonious trees were discovered (121 steps; CI = 0.80; RI = 0.88). In each of these 10 trees, *Cienfuegosia tripartita* occupied position **A**, but the placement of *Thespesia thespesioides* was unstable, appearing either at position **i** or **ii** ("**A-ii**"). The ten most-parsimonious trees represent two sets of five trees (**A-i** and **A-ii**) corresponding to five different arrangements in *Gossypium*, shown as G1, G3, G4, G5, and G6 in the inset of Fig. 3.

In the fourth and final analysis, gaps were coded as presence/absence characters, but the gapped positions were excluded from the data matrix. As with the third analysis, 10 shortest trees were found (120 steps; CI = 0.80; RI = 0.88), corresponding to five each for topologies **A–i** and **A–ii**. For each of these two topologies, five different arrangements were observed in *Gossypium* (G1, G2, G3, G5, and G6).

Three of the five gaps introduced into the aligned sequences (gaps **a**, **d**, and **e**, Fig. 2) appear to represent autapomorphic indels. Regardless of the topology chosen in Fig. 3, gap **a** is explained as a two-codon insertion in the lineage leading to *Cienfuegosia tripartita*. Gaps **d** and **e** are ambiguous with respect to polarization in that they each occur in only one of the two outgroup taxa (gap **d** in *Hibiscus costatus*; gap **e** in *Anotea flavida*). Gap **c** seems most-parsimoniously interpreted as a synapomorphic two-codon deletion for all genera in the tribe except *Cienfuegosia*. Indeed, trees in which *Cienfuegosia tripartita* is basal within the tribe (**A-i** and **A-ii** in Fig. 3) require only a single step for this character, whereas two steps (a basal deletion in the ingroup followed by an insertion in the *Cienfuegosia tripartita* lineage) are required in trees where *Cienfuegosia tripartita* occupies positions **B** or **C**. Thus, this single gap, considered in isolation of other data, constitutes evidence for a basal phylogenetic position of *Cienfuegosia* within the tribe.

The fifth and final gap (gap **b**) is homoplasious regardless of the topology selected for mapping the character. In trees where *Thespesia thespesioides* is at position **ii**, fewer steps are required than in trees where

this taxon is at position **i**. If this observation reflects an actual most-parsimonious history, then gap **b** is interpreted as a two-codon insertion in the clade consisting of *Hampea* and *Lebronnecia* and an independent insertion in the lineage leading to *Gossypium*, *Gossypioides*, and *Kokia*. In addition, it appears that there have been several reversals within *Gossypium*, specifically in the ancestors of the A-genome, B-genome, and E-genome taxa. In this respect, each of the six arrangements in *Gossypium* (G1 through G6, Fig. 3) require three steps for gap **b**. We infer that the codon positions represented by gap **b** are evolutionarily highly labile within the tribe.

In summary, the gap information provides supporting evidence, albeit somewhat scanty, for the trees in which *Cienfuegosia tripartita* occupies position **A** and *Thespesia thespesioides* occupies position **ii**. Thus, tree **A-ii**, in which *Cienfuegosia* is basal and *T. thespesioides* is sister to *Hampea/ Lebronnecia*, is the preferred resolution on the basis of information presently available.

Notwithstanding the several ambiguities, all topologies recovered, regardless of gap treatment, agree with respect to a number of cladistic events. The most noteworthy of these are: (1) *Gossypium* is monophyletic; (2) *Kokia* is sister to *Gossypioides*; (3) *Gossypium* shares a most recent common ancestor with the clade consisting of *Kokia* and *Gossypioides*, with these three genera composing one of two monophyletic generic groupings in the tribe; (4) *Hampea*, *Lebronnecia*, and *Thespesia* are suggested to be a second monophyletic grouping of genera in the tribe; and (5) although sampling was limited, *Thespesia* appears not to be monophyletic.

With respect to relationships within the genus *Gossypium*, as many as six alternatives were revealed by the various analyses. All of these indicated that the clade consisting of A-,AD-, and F-genome species is monophyletic, and that the B- and C-genome groups represent sister taxa. Other relationships within the genus are ambiguous.

We conducted distance-based phylogenetic analyses using neighbor-joining trees and a matrix of Kimura two-parameter distances derived from the *ndhF* data. Regardless of how gaps were treated in these analyses, a topology was obtained (Fig. 3, bottom) that exhibits many similarities to trees recovered from the character-based analyses. In particular, all five

conclusions enumerated for the character-based analyses are supported by the neighbor-joining results. In addition, the depiction shown highlights the existence of some short interior branches, echoing the phylogenetic instability observed for the placement of *Cienfuegosia* and for relationships among species groups in *Gossypium*.

Variation in ITS. We previously presented ITS sequences from 24 species of *Gossypium*, mostly from A-, D-, and AD-genome species (Wendel et al. 1995a, 1995b). In this work absolute lengths of the spacer region (ITS1 + 5.8S gene + ITS2) varied from 669 bp to 685 bp, with an aligned length of 687. In the present study, 39 new sequences were generated from 33 species not represented in the earlier studies and including all genera in the tribe except for *Cephalohibiscus* (Table 1). Incorporating these new sequences lengthened the aligned data base to 729 bp. Excluding gaps, the entire ITS region varied from 588 nucleotides in *Hampea appendiculata* to 689 bp in *Gossypium londonderriense*, with most sequences being 682-684 bp in length. Most alignment gaps were 1 bp in length: 22 in ITS1; 1 in the 5.8S gene; and 9 in ITS2. Five additional gaps ranged from 2 - 4 bp in length. The final data matrix consisted of 64 rows (sequences) and 729 columns (characters). Alignment of ITS sequences from all ingroup taxa was straightforward, but was problematic when ITS sequences from the two outgroup taxa were included. This experience suggests that homologies in some parts of the sequence may be sufficiently uncertain to preclude the use of this region for family-wide studies in the Malvaceae.

Aligning the ITS1 region resulted in a matrix 321 nucleotides in length, with absolute lengths ranging from 218 bp (in *Hampea appendiculata*) to 297 bp (in *Thespesia populnea*). Sequences from most species of *Gossypium*, *Lebronnecia kokioides*, *Gossypioides kirkii*, *Kokia*, and *Anotea flavida* were near the upper end of this range (293-295 bp), whereas those from the remaining taxa in the Gossypieae and from the outgroup taxon *Hibiscus costatus* were considerably shorter (264-268 bp). Most length variation in ITS1 was due to 41 and 36 bp gaps near the 5' end of the sequence from *Hampea appendiculata* (Fig. 4, top panel), and to a 36-bp gap in *Cienfuegosia* and *Thespesia* sect. *Lampas* (Fig. 4, middle panel).

Sequencing of the 5.8S gene was incomplete for some taxa, in that approximately 40 bp surrounding the primer annealing sites was not

sequenced. Absolute lengths for the complete sequences varied from 164-168 nucleotides, with an aligned length of 169. One four-bp gap was due to an apparent insertion in *Lebronnecia kokioides* and a single one-bp gap reflects an additional putative insertion in *Hibiscus costatus*. Apart from these anomalous sequences, all other 5.8S genes fully sequenced were 164 bp long.

Absolute lengths for the ITS2 region varied from 206 bp in *Hampea appendiculata* to 229 bp in *Kokia*, with most sequences being 224 - 225 nucleotides in length. Most of the length variation was due to a 15 bp gap near the beginning of ITS2 in *Lebronnecia kokioides*, *Thespesia* sect. *Lampas*, and *Hampea appendiculata* (Fig. 4, bottom panel) and a second 15 bp gap at the end of ITS2 in several species of *Gossypium* (Wendel et al. 1995a). Aligned sequences were 239 nucleotides in length.

A high degree of variation exists in the spacer regions. Excluding outgroup sequences, approximately one-third of all positions exhibited nucleotide polymorphisms in ITS1 (108/321 = 33.6%) with a slightly higher percentage in ITS2 (104/239 = 43.5%). Of this variation, 73 (22.7%) and 67 (28.0%) of the characters in ITS1 and ITS2, respectively, were potentially phylogenetically informative. In contrast to this high variability, the 5.8S gene was strongly conserved in sequence and provided only five potentially phylogenetically informative characters. For the entire ITS region, 30.7% (224) of the 729 aligned nucleotide positions were variable, of which two-thirds (145) were potentially phylogenetically informative.

Phylogenetic analysis of ITS sequences. Phylogenetic analysis yielded 4212 most-parsimonious trees, each 706 steps long with consistency and retention indices of 0.57 and 0.79, respectively. As was the case with analysis of the *ndhF* data, the phylogenetic position of *Cienfuegosia* is ambiguous with respect to the ITS data, although in this case only two different placements appeared in the suite of most-parsimonious trees. In approximately 85% of the trees, *Cienfuegosia* is basal within the tribe (at position "**a**" in Fig. 5; topology **a** hereafter). In the remaining 15% of the trees, *Cienfuegosia* occupies position **b** in Fig. 5 (topology **b** hereafter), still a relatively early divergence within the tribe, but subsequent to the separation of the clade consisting of *Hampea*, *Lebronnecia*, and *Thespesia* sect. *Lampas*. Although the position of

Cienfuegosia at **a** in Fig. 5 was also recovered in the *ndhF*-based tree (Fig. 2), the other position for *Cienfuegosia* in Fig. 5 (position **b**) was only observed in ITS-based trees.

In addition to uncertainty regarding the position of *Cienfuegosia*, other cladistic patterns resulting from the ITS data parallel those derived from the *ndhF* analyses. Specific points of agreement at the generic level include: (1) *Gossypium* is monophyletic and derived within the tribe; (2) *Kokia* is sister to *Gossypioides*; (3) *Gossypium* shares a most recent common ancestor with the clade consisting of *Kokia* and *Gossypioides*, with these three genera constituting a monophyletic generic grouping; (4) *Thespesia* appears not to be monophyletic; and (5) *Hampea*, *Lebronnecia*, and *Thespesia* sect. *Lampas* also comprise a monophyletic generic grouping. In contrast to this broad agreement between the *ndhF*- and ITS-derived phylogenies, there are two notable differences. The first regards the position of *Cienfuegosia*, as discussed. Second, in *ndhF*-based trees, *T. populnea* is sister to the *Hampea/Lebronnecia/T. lampas* clade, but in trees based on ITS sequences, *T. populnea* is sister to the *Gossypium/Kokia/ Gossypioides* clade.

Gossypium was the most intensively sampled genus, with 52 sequences generated. Four alternatives were evident in the 4212 most-parsimonious trees for relationships among the major genomic groups: in topology **a**, E-genome species were located at one of the positions denoted by **i** through **iv** in Fig. 5, whereas in topology **b**, E-genome species resolved at positions **i** or **iii** or **iv**. All trees supported an early divergence of Asian-African A-, B-, and F-genome species followed by separation of the New World, D-genome species from the clade consisting of Australian, C-genome species and African, E-genome species. Of the five AD-genome species, four occurred within the D-genome clade whereas the other (*G. mustelinum*) was nested within the A-genome clade, consistent with earlier results (Wendel et al. 1995a).

Combined analysis in the Gossypieae. Two strictly parallel data sets, one from the plastid genome (*ndhF*) and one from the nuclear genome (ITS), were generated by pruning the accessions from the ITS data set for which we had no *ndhF* sequences. For *ndhF*, the strict consensus of the 15 shortest trees found when gaps were coded as missing is shown in panel A

of Fig. 6. We chose this consensus over those derived from other gap codings because results from each of the four analyses were so similar (Fig. 2). As shown in Fig. 6, two areas of ambiguity are identified in the consensus, resulting in polytomies for the basal divergence in the tribe (drawn as a trichotomy for *Cienfuegosia*, *Lebronnecia/Thespesia/Hampea*, and *Gossypioides/Kokia/Gossypium*), and for relationships among genomic groups in *Gossypium*.

For ITS, four most-parsimonious trees were found, each 487 steps long with consistency and retention indices of 0.61 and 0.60, respectively (uninformative characters excluded). The strict consensus derived from these trees is shown as panel B in Fig. 6. As with the *ndhF* data, one of the two trichotomies in this strict consensus reflects ambiguity concerning the basal divergence in the tribe, although in this case *Thespesia populnea* is resolved as sister to the *Gossypioides/Kokia/Gossypium* clade rather than as a part of a *Lebronnecia/Thespesia thespesioides/Hampea* clade. The other trichotomy is among taxa in this latter clade.

A combined data set was formed by merging the *ndhF* and ITS data. Phylogenetic analysis of this combined matrix led to the recovery of four shortest trees (L = 618, excluding uninformative characters; CI = 0.63; RI = 0.65), the strict consensus of which is shown in panel C of Fig. 6. As was the case for trees obtained from analyses of the separate data sets, a basal polytomy is evident, reflecting ambiguity in the divergence order for *Cienfuegosia*, *Lebronnecia/Thespesia thespesioides/Hampea*, *Thespesia populnea* and *Gossypioides/Kokia/Gossypium*. This polytomy actually reflects two separate ambiguities, one concerning the basal clade (*Cienfuegosia* or *Lebronnecia/Thespesia thespesioides/Hampea*) and the other the position of *Thespesia populnea* (sister to *Gossypioides/Kokia/Gossypium* or *Lebronnecia/Thespesia thespesioides/Hampea*). In addition, the trichotomy for *Lebronnecia/Thespesia thespesioides/Hampea* found in the ITS-derived consensus is retained in the consensus derived from the combined analysis. A third ambiguity is for relationships among genomic groups in *Gossypium*, as shown.

A number of phylogenetic inferences are supported by both separate and combined analyses. These include: (1) evidence, albeit tentative due to the limited sampling outside the tribe, for monophyly of the Gossypieae; (2)

the close relationship among *Lebronnecia*, *Thespesia thespesioides* and *Hampea*; (3) the sister-taxon relationship between *Gossypioides* and *Kokia*; (4) the sister-taxon relationship between the *Gossypioides/Kokia* clade and *Gossypium*; and (5) the monophyly of *Gossypium*. Branch support, as measured by minimum character number and decay indices, is high for most of these inferences (Fig. 6), thereby suggesting that they are robust.

Congruence and incongruence in the Gossypieae. Many areas of agreement are apparent between the *ndhF* and ITS topologies. Character incongruence was quantified using the incongruence indices I_{MF} and I_M, as summarized in Table 3. These are modest estimates in comparison to other data sets (e.g., Kim and Jansen 1994; Omland 1994).

As discussed above under *Data Analysis*, previous phylogenetic work has documented a number of examples of reticulation in the genus *Gossypium*. To explore the effect of these obvious sources of character incongruence on the incongruence indices, we recalculated I_{MF} and I_M after excluding the reticulate taxa (Table 3). These changes had no effect on the *ndhF*-, ITS-, and combined phylogenies, but incongruence indices sharply decreased to 0.026 and 0.095 for I_{MF} and I_M, respectively. Furthermore, when the B-genome species *Gossypium anomalum* was excluded along with the known or suspected reticulate taxa (due to the different placement in *ndhF* and ITS trees; see Fig. 6 trees A and B), I_{MF} and I_M were decreased to very low values, viz., 0.011 and 0.033, respectively.

In addition to quantifying character incongruence, we used WSR tests to examine specific points of incongruence between the data sets. These tests are summarized in Table 4, for nodes illustrated in Fig. 6. When all taxa are included, each data set rejects the consensus tree derived from the alternative data set, and the *ndhF* data set also rejects the strict consensus tree from combined data set. When the *a priori* sources of incongruence (i.e., reticulate taxa) were removed, however, neither data set rejects the topology reflected in the strict consensus derived from the alternative data set. This effect is most evident when *Gossypium anomalum* was excluded as well (Table 4, part 1C).

To obtain information on the specific nodes causing incongruence, tests of individual branch points were conducted and compared for analyses done with and without the reticulate taxa. As shown in Table 4, both the

ITS and *ndhF* data sets statistically reject (by WSR tests) resolutions in the alternative trees only for some nodes in *Gossypium* and only when all taxa are included in the analysis. When reticulate taxa are excluded, the only significant value ($p = 0.02$) is obtained when ITS data are used to test the monophyly of (B + C)-genome *Gossypium* indicated in the *ndhF*-derived consensus. When the B-genome species, *G. anomalum*, is also excluded from the data sets, the statistical incongruence between data sets disappears. These analyses demonstrate that (1) significant incongruence between the ITS and *ndhF* data sets arises exclusively from character conflict in *Gossypium* rather than from conflict in other genera (cf. Table 4); and (2) that the taxonomic sources of incongruence may be identified by phylogenetic and statistical analysis.

Combined analysis in the Genus Gossypium. Two parallel data sets were available for *Gossypium*, the ITS data reported here and the cpDNA restriction site data of Wendel and Albert (1992). Reanalysis of the latter data yielded the same four shortest trees reported earlier, each 161 steps long (excluding uninformative characters) and with consistency and retention indices of 0.84 and 0.81, respectively. In all four trees, species belonging to the same cytogenetic (genomic) group are monophyletic and most of these monophyletic groups have high decay indices and character support. The cpDNA data also support the notion that the C-genome clade diverged first from the remainder of the genus. In general, phylogenetic trees derived from cpDNA restriction site data agree well with traditional taxonomic views at the sectional or subsectional levels (Fryxell 1979, 1992; as discussed by Wendel and Albert 1992). Notable differences among the trees concern the positions of B- and F-genome species. With respect to the former, the B-genome clade resolves either as the sister taxon to the D-genome clade or forms a trichotomy with the D-genome clade and the clade consisting of A-, AD-, E-, and F-genome species. Ambiguity involving the lone F-genome species (*G. longicalyx*) is due to its placement as the sister taxon to either the A-genome + AD-genome clade or to the E-genome clade. Because of these instabilities, there is incomplete resolution in the strict consensus (panel A of Fig. 7).

Phylogenetic analysis of ITS sequences from the same taxa as those included in the cpDNA restriction site data set resulted in 90 most-

parsimonious trees (L = 222, excluding uninformative characters; CI = 0.63; RI = 0.53), the strict consensus of which is shown in panel B of Fig. 7. As with the cpDNA-derived trees, all genomic groups appear as monophyletic, once *G. gossypioides* (Wendel et al. 1995a) and the allopolyploids (Wendel et al. 1995b) have been justifiably removed from consideration. Most of the differences among the 90 shortest trees were alternative resolutions within several terminal species groups, which are, accordingly, shown as unresolved in the strict consensus. In addition, relationships among A-, B-, and F-genome species groups were ambiguous, as were branch orders for C- and E-genome species groups.

There are many areas of agreement between the ITS- and cpDNA-based phylogenies, and several striking incongruences as well. Some of these were expected; for example, allopolyploid *Gossypium* species contain an A-genome cytoplasm (Wendel 1989), and hence, the position of the allopolyploid species is expected to differ in trees based on maternally inherited (plastid) and biparentally inherited (nuclear) markers. Other differences, however, are not as readily explained, such as differences in the placement of the B- and E-genome species and relationships among species groups within the C- and D-genome clades. When the ITS and cpDNA data sets are combined, phylogenetic analysis yields 10 most-parsimonious trees (L = 468; CI = 0.59; RI = 0.85). The strict consensus of these (panel C, Fig. 7) demonstrates that the differences between the two sources of data conflict to such an extent that there is a near-complete loss of resolution among genomic groups. All ten trees agree that the C-genome clade is basal to the rest of the genus, but no other bifurcations among genome groups are supported.

When all taxa known or suspected to have reticulate histories were excluded from the data sets, there were no significant topological changes in either the cpDNA- or ITS-based phylogenies regarding relationships among the remaining taxa. However, the numbers of most-parsimonious trees changed. With the cpDNA restriction site matrix, eight most-parsimonious trees were discovered instead of four (L = 137; CI = 0.83; RI = 0.96); a strict consensus of these trees is shown in panel A of Fig. 8. In constrast to the increase in number of shortest trees detected when the cpDNA restriction

site data were analyzed, the ITS data yielded only 12 shortest trees (L = 179; CI = 0.68; RI = 0.85) when reticulate taxa were excluded, 78 fewer than when they were included. The strict consensus of these 12 trees is shown in panel B of Fig. 8.

Not surprisingly, when the ITS and cpDNA data were combined and reticulate taxa were excluded, phylogenetic analysis led to greatly enhanced resolution compared to the trees obtained when reticulate taxa were included (Fig. 7, panel C vs. Fig. 8, panel C). As shown in Fig. 8, only one shortest tree (L = 342) was found, with a fully resolved phylogeny of genomic groups in the genus, and with somewhat higher consistency and retention indices than when reticulate taxa were included (CI = 0.70 vs. 0.59; RI = 0.89 vs. 0.85). Several noteworthy phylogenetic inferences may be drawn from this tree: the monophyly of each of the diploid genome groups A through F, the basal position of the C-genome clade, and the non-monophyly of the African species (A-, B-, E-, and F-genomes) in their entirety but the collective monophyly of three (A, B, F) of the constituent groups. In addition, branch support numbers and decay indices for most genomic groups are relatively high, although we note that support is weak (decay index = 1) for the bifurcation of the E- and the (A + F + B + D)-genome clades, as well as for the clade uniting the D- with the (A + F + B)-genomes. The combined data set does not strongly support this particular resolution of basal relationships among genomic groups, despite the complete resolution and the high retention index.

Congruence and incongruence in Gossypium. The I_{MF} and I_M show that the cpDNA and ITS data sets are highly incongruent (Table 3). Eliminating reticulate taxa from the data sets resulted in a decrease in the incongruence estimates (to 19.6% and 53.7% for I_{MF} and I_M, respectively), although these values indicate that character conflict remains in the pruned data sets.

In an effort to isolate the nodes responsible for the conflict, WSR tests were conducted, as summarized in Table 4 for the nodes illustrated by roman numerals in Figs. 7 and 8. As expected, most tests returned low (significant) probability values when all taxa were included. When the nine known or suspected reticulate taxa were excluded, most of the incongruence dissipated, although the WSR tests indicate that the cpDNA and ITS data

sets retain significant incongruence. Specifically, the cpDNA data strongly reject the ITS strict consensus tree and weakly reject three other relationships shown in the ITS consensus tree: monophyly for the (C + E + D)-genomes; basal divergence of the (A + B + F)-genome clade; and the sister taxon status for Sections *Sturtia* and *Hibiscoidea* in the C-genome clade. The cpDNA data fail to reject, but just barely ($p = 0.06$), the combined tree. Similarly, the ITS data are statistically incongruent with the cpDNA strict consensus tree; the strict consensus derived from the combined data set; and monophyly of the (A + E + F)-genomes, as observed in the cpDNA-based tree. We note that the ITS data fail to reject monophyly for the (A + E + F + B + D)-genome clade and also fail to reject the basal separation of the C-genome clade, both of which represent strongly supported inferences in the cpDNA-based consensus tree.

DISCUSSION

Congruence and incongruence in the Gossypieae. In this study, two data sets were used to explore cladistic relationships in the Gossypieae, one from the plastid genome (*ndhF*; Fig. 2) and one from the nuclear genome (ITS; Fig. 5). A comparison of the phylogenies inferred from these two data sets shows that they are largely congruent, even with respect to the ambiguity of resolution near the root of the trees. When the data sets are pruned so that only taxa in common to both are included in the analyses, only two areas of incongruence are evident (Fig. 6), one for relationships among genomic groups in *Gossypium*, and the other concerning the placement of *Thespesia populnea* (either as sister to the *Gossypioides/Kokia/Gossypium* clade or to the *Lebronnecia/Thespesia thespesioides/Hampea* clade).

Even though the two data sources returned nearly congruent topologies, and hence low incongruence indices (Table 3), the question arises as to how to evaluate and analytically treat the incongruence that is observed. Because incongruence may be either "real", that is, reflective of one or more underlying biological phenomena, or "spurious", meaning that it arises from insufficient data or some other unknown artifact, the strategy we adopted was to evaluate whether there is sufficient evidence to conclude that the observed incongruence is real, prior to seeking underlying biological

explanations or making decisions about pruning particular taxa for combined analyses. To accomplish this, we used routine methods for evaluating the robustness of clades (character support, decay indices) in addition to quantitative methods for measuring incongruence and testing its statistical significance. This process allowed the primary contributors to overall incongruence to be identified and, if desired, pruned prior to conducting a combined analysis.

With respect to generic relationships within the Gossypieae, this strategy demonstrated that the different positions occupied by *Thespesia populnea* in the *ndhF* (Fig. 6A) and ITS (Fig. 6B) trees are statistically equivalent, as judged by the results from the Wilcoxon Signed Rank tests (Table 4). Specifically, ITS data set fail to reject the *Lebronnecia/Hampea/Thespesia* monophyly observed in the *ndhF* tree (node i in part 1 of table 4; $p = 0.37$), and the *ndhF* data fail to reject the arrangement observed in the ITS tree (node iv in part 1 of table 4; $p = 0.25$). Thus, there is insufficient evidence to conclude that the placement of *Thespesia populnea* differs significantly in the two phylogenies, and accordingly, we conclude that the two data sets do not disagree.

The second area of potential incongruence between the two data sets is for genomic relationships within *Gossypium*. In this case, WSR tests return a number of significant results when each data set is used in combination with constraint trees derived from the other data set (Table 4). Among the several possible causes for this is the inclusion of reticulate taxa in the analyses. As discussed above, a number of taxa included in the study were expected to resolve differently in phylogenetic analyses based on maternally inherited and biparentally inherited characters, because of allopolyploidy (Wendel 1989), recombinant histories (Wendel and Albert 1992) or interlocus concerted evolution (Wendel et al. 1995b). When these "reticulate taxa" (in this case *G. cunninghamii*, *G. hirsutum*, and *G. barbadense*) are removed from the data sets, the inferred phylogenies remain the same (except for the excluded taxa) for both *ndhF* and ITS, but the total amount of incongruence between the two data sets is reduced sharply to relatively low values (0.229 to 0.096 and 0.079 to 0.026 for I_M and I_{MF}, respectively; Table 3). Moreover, the incongruence between *ndhF* and ITS data sets for specific nodes, as measured by WSR tests, is nearly

eliminated. These results confirm that the previously described organismal or molecular reticulation is the primary cause of incongruence between the *ndhF* and ITS data sets.

Although incongruence indices are reduced to low levels when the reticulate taxa are excluded, not all incongruence disappears. In addition, WSR tests show that the *ndhF* node showing (B + C)-genome monophyly is rejected ($p = 0.02$) by the ITS data set. In this case, the offending taxon appears to be the B-genome species, *G. anomalum*. When *G. anomalum* was excluded from the analyses, both data sets are nearly perfectly congruent (I_M and $I_{MF} = 0.033$ and 0.011, respectively; Table 3). Although WSR tests were not performed for all nodes of trees inferred from the *ndhF* and ITS data sets, no significant values were obtained for any of those nodes tested when all reticulate taxa and *G. anomalum* were excluded.

In summary, the *ndhF* and ITS data sets are highly congruent with respect to relationships among genera, and are nearly congruent for *Gossypium* once reticulate taxa are excluded. This topological agreement suggests that a combined analysis of a single pooled data set will lead to the best estimate of phylogeny given the available data (Bull et al. 1993). Because *Gossypium* is discussed in detail below and because generic-level relationships are not affected by the inclusion or exclusion of reticulate taxa in *Gossypium* (not shown), the topology obtained from combined analysis is captured in the strict consensus shown in Fig. 6C. This topology retains considerable resolution and reflects what we believe are six noteworthy phylogenetic inferences.

First, insofar as it has been evaluated, the Gossypieae is monophyletic. The molecular data, therefore, support and augment the morphological synapomorphies of punctae or pigment glands that synthesize terpenoid aldehydes (collectively referred to as "gossypol"), and the cotyledonary and seed coat characters described by Fryxell (1968).

Second, even though the *ndhF*, ITS and combined data sets resulted in ambiguity with respect to the earliest divergence in the tribe (Fig. 6), we suggest that the lineage now represented by *Cienfuegosia* represents the descendants of one of the two basalmost branches. The only phylogenetic evidence in support of this suggestion, as far as we know, is the presence of a two-codon insertion in *ndhF* (gap c; Fig. 3) found in *Cienfuegosia* and the

outgroup taxa but not in other ingroup taxa. If this gap is included in phylogenetic analysis, the basal polytomy disappears and *Cienfuegosia* becomes sister to the remainder of the tribe. In addition to this phylogenetic character, we note that *Cienfuegosia* is phenetically "...the genus most isolated from the remainder of the genera [in the Gossypieae]..." (Fryxell 1979, p. 211) and that its members have chromosome numbers (2n = 20, 22) that are not found elsewhere among related genera (2n = 24, 26).

Third, all analyses indicate a close relationship among *Hampea*, *Lebronnecia*, and *Thespesia* sect. *Lampas*, although this relationship is unresolved in the combined analysis (Fig. 6). We are aware of no previous suggestion that these three taxa are more closely related to each other than they are to other genera in the tribe, and we were unable to discern a morphological synapomorphy for the molecular clade.

Fourth, although sampling was minimal, *Thespesia* is not monophyletic in our analyses. Instead, *Thespesia* appears as either paraphyletic to *Hampea* and *Lebronnecia* (*ndhF* phylogeny) or polyphyletic (ITS and combined phylogenies). Our results reinforce the previous recognition of "...a discontinuity at the sectional level separating...section *Thespesia* and section *Lampas*..." (Fryxell 1979, p. 100).

Fifth, and remarkable given their geographic ranges, the East African - Madagascan genus *Gossypioides* is sister to the Hawaiian endemic genus *Kokia*. This inference appears to be robust, as decay indices and character support numbers for this clade are high in all analyses. Although we are aware of no unifying morphological feature for this clade, the two genera are close phenetically (Fryxell 1979, p. 214) and they share a somatic chromosome number of 24. Chromosome numbers have been difficult to polarize within the tribe due to the wide variation in numbers reported in the Hibisceae and Malvavisceae. Given the robust observation of a *Gossypioides-Kokia* clade, however, it seems probable that 2n = 24 is a state derived from a single aneuploid reduction in an ancestor with 26 chromosomes, as presently found in *Gossypium*, *Hampea*, and most of *Thespesia*. If this scenario reflects actual history, a synapomorphic loss of one pair of chromosomes would have occurred in the lineage leading to *Kokia* and *Gossypioides* after the divergence of the branch leading to modern *Gossypium*. The most parsimonious mechanisms are either the

fusion of two nonhomologous chromosomes or the loss of a chromosome pair from the genome. Tantalizing evidence in favor of the fusion hypothesis comes from Hutchinson (1943), who observed that one chromosome pair of *Gossypioides brevilanatum* Hochr. (as *Gossypium brevilanatum*) is unusually long compared with all other chromosomes and that this observation was a "marked feature of every plate examined." Hutchinson (1943) also remarked in a footnote that successful grafts could be made between *Kokia rockii* [=*Kokia drynarioides*] and *Gossypioides kirkii*. In the context of our cpDNA and ITS data, this observation seems more meaningful.

Finally, the genus *Gossypium* is monophyletic and is sister to the *Gossypioides/Kokia* clade, a relationship that has not previously been suspected. Support is high for the monophyly of *Gossypium* in both separate and combined analyses, and for the sister relationship with *Gossypioides/Kokia* in the ITS and combined analysis (Fig. 6). It is therefore disconcerting that not a single, morphological synapomorphy for *Gossypium* has been identified that is not also homoplasious within the tribe. For example, although this triad of genera is weakly supported by the presence of foliose (vs. subulate) epicalyx bracts (Fryxell 1979), the non-foliose condition is also common within *Gossypium*.

Origin and radiation of the Gossypieae. Palynological records indicate that the oldest malvacean pollen type is from the Eocene (38 - 45 million years before the present, mybp) in South America and Australia and from the Oligocene (25 - 38 mybp) in Africa (Muller 1981, 1984; Macphail and Truswell 1989). The Malvaceae thus probably originated during the first third of the Tertiary and achieved a world-wide distribution approximately 30 mybp. Beyond this, there is little in the fossil record that directly reflects upon the origin or diversification of the Gossypieae. The *ndhF* sequence data provide an alternative source of information, in that divergence levels may be used to estimate divergence times, given a molecular clock. There are a number of important caveats and limitations to the use of clocks based on sequence data (Hillis et al. 1996), including substitution rate heterogeneity among lineages within a study, uncertainties regarding clock calibration, and unknown, but presumably large, estimation errors. Despite these limitations, we view the *ndhF* sequence data as providing approximate divergence time estimates, which, if in agreement

with the palynological record, may contribute to our understanding of the history of divergence and dispersal in the Gossypieae.

To apply a clock based on *ndhF* sequences, we first evaluated whether substitution rates were homogeneous within the tribe. Two types of tests were conducted using the methods of Takezaki et al. (1995): "two-cluster" tests, which evaluate whether two clusters created by a branch point in a given tree have different nucleotide substitution rates, and "branch-length" tests, which examine whether branch length between the tree root and a tip deviates from the average branch length. In these tests, we used the nucleotide substitution model of Kimura (1980; thereafter called K2P), and *Hibiscus costatus* and *Anotea flavida* were the outgroup cluster. The two-cluster tests indicated that *ndhF* substitution rates were constant across all taxa (data not shown), but branch-length tests revealed that the sequence from *Gossypium cunninghamii* evolved significantly ($p < 0.05$) faster than average, and sequences from *Thespesia populnea* and *T. thespesioides* evolved significantly ($p < 0.01$ and $p < 0.05$, respectively) slower than average. Tajima's (1993) method similarly yielded test results indicating mostly homogeneous substitution rates. Apart from the few sequences with slightly accelerated or decelerated rates, the rate tests establish that, in general, *ndhF* sequences in members of the Gossypieae have homogeneous substitution rates.

A more problematic aspect of divergence time estimation is clock calibration. The absolute rate of nucleotide substitution for *ndhF* is not known for any pair of angiosperm species, and even if it were, the common observation of rate heterogeneity across lineages would render the use of that rate suspect. As an alternative, we used an average rate for single copy plastome genes that we estimated from Table IV of Palmer (1991, p. 39) to be 5×10^{-10} nucleotide substitutions per site per year. This estimate obviously is associated with an error, although its magnitude is unknown.

Using this estimate of absolute rate and the *ndhF* sequence divergence estimates, we calculated approximate divergence times for various divergences within the tribe and within *Gossypium* (Table 5). Within the tribe, the mean sequence divergence between the two basal clades (*Cienfuegosia* and other genera) is 1.89%, which translates into an initial divergence of 19 mybp. This estimate, if accurate, functions as the lower

bound on the age of the tribe, and is consistent with the palynological evidence cited above. We also estimate that *Lebronnecia*, *Thespesia* and *Hampea* separated from the rest of tribe approximately 15 mybp and from each other 7 - 10 mybp, and that *Gossypium* diverged from its sister clade (*Kokia* and *Gossypioides*) approximately 12.5 mybp. Within *Gossypium*, divergences among the major genomic groups appear to have been closely spaced (Table 5), with the major lineages having been established approximately 11.5 mybp, shortly after divergence from the *Kokia-Gossypioides* clade. Divergence between *Kokia* and *Gossypioides* bears special mention, as it concerns a biogeographically novel cladistic relationship between endemics from Hawaii and Africa/Madagascar. The amount of *ndhF* sequence divergence between *Kokia* and *Gossypioides* (K2P = 0.29%) is less than half that of any other intergeneric comparison in the tribe, and is even lower than that observed between most inter-genomic comparisons within *Gossypium*. The *ndhF* data therefore suggest that the separation of the lineages that gave rise to *Kokia* and *Gossypioides* occurred in the Pliocene, about 3 mybp. This implicates trans-oceanic dispersal as a factor in the evolution of one or both genera, either directly between the Hawaiian archipelago and Africa/Madagascar, or via extinct intermediates on these or other oceanic islands. In this respect, the *Kokia-Gossypioides* floristic relationship represents only the latest in a series of examples of long-distance, salt-water dispersal in the Gossypieae (Stephens 1966; Fryxell 1979; DeJoode and Wendel 1992; Wendel and Albert 1992). To the extent that the estimates provided for other divergences are accurate (Table 5), they similarly implicate oceanic dispersal as a major factor in the evolution of the tribe.

Congruence and incongruence in Gossypium. The cpDNA restriction site data of Wendel and Albert (1992) and the ITS sequence data reported were both used to estimate phylogeny within the cotton genus. When analyzed either separately or as total evidence, these data sets lead to phylogenies that agree in several important respects (Figs. 7 and 8). Most noteworthy is the monophyly of each of the diploid cytogenetic groups, a cladistic pattern that is concordant with geographic and taxonomic alignments (as discussed by Wendel and Albert 1992). In addition to

genomic monophyly, the cpDNA- and ITS-based phylogenies show congruence with respect to relationships among some species within clades.

Perhaps of more interest than the agreement, however, is the high level of incongruence between the cpDNA and ITS data sets (I_M = 0.736; I_{MF} = 0.438; Table 3). Much of this conflict arises from inclusion of the "reticulate" species, *G. bickii* (Wendel et al. 1991), *G. cunninghamii* (Wendel and Albert 1992), *G. aridum* (DeJoode 1992; Wendel and Albert 1992), *G. gossypioides* (Wendel et al. 1995a), and the allopolyploids (Wendel 1989; Wendel et al. 1995b), each of which occupied conflicting positions in the cpDNA- and ITS-derived trees (Fig. 7). The effect of these taxa on the total incongruence is clear; when *G. bickii*, *G. cunninghamii*, *G. aridum*, *G. gossypioides* and the allopolyploids are sequentially excluded from the data sets, total incongruence decreases noticeably (I_M is reduced from 0.736 to 0.712, 0.693, 0.675, 0.658, and 0.537 respectively).

This exercise demonstrates that, as expected, a portion of the incongruence is caused by each of the reticulate species or species groups in the phylogenetic analysis. However, both the amount of incongruence remaining once the reticulate taxa have been removed, and inspection of the pruned topologies (Fig. 8), show that considerable conflict remains between the cpDNA and ITS data. Some disagreements involve relationships among genomic groups and reflect conflict in branching order for relatively early divergences in the evolution of *Gossypium*. Three examples, in particular, possibly account for the incongruence at this level: (1) the basal separation in the genus; (2) the placement of subsection *Pseudopambak* (E-genome); and (3) the position of subsection *Anomala* (B-genome). In addition to these conflicts involving deep divergences, there are disagreements between trees with respect to resolution of more recently diverged taxa that are not suspected to have reticulate ancestries. This is most apparent for relationships among subsections within the D-genome and C-genome clades. We discuss and evaluate each of these incongruent resolutions in the following paragraphs, using for reference the topologies for which reticulate taxa have been excluded (Fig. 8).

Perhaps the most striking difference between the cpDNA and ITS phylogenies concerns the first dichotomy in the genus. In the cpDNA tree, the earliest separation was between a monophyletic C-genome clade and the

remainder of the genus, but in the ITS tree, a (C + E + D)-genome clade appears as sister to an (A + B + F)-genome clade. The latter resolution, however, is not strongly supported. In fact, the cpDNA data are more internally consistent than the ITS data, as evidenced by the higher retention index (0.96 vs. 0.85). In addition, a basal separation of the C-genome clade has more character support and a higher decay index than the alternative (Fig. 8). It is not surprising, therefore, that with respect to this single issue, the cpDNA topology is retained in the combined analysis. This differential level of support between the two data sets also helps to explain the results of the WSR tests (Table 4): the ITS data set does not reject the C-genome clade as basal ($p = 0.15$), but the cpDNA restriction site data statistically rejects the alternative shown in the ITS tree ($p = 0.02$). The weight of the evidence thus appears to favor the hypothesis that the C-genome species represent the descendants of one branch from the earliest split in the genus.

A similar set of indicators bears on the incongruence observed for the placement of the E-genome clade, which is included with other African groups in the cpDNA tree but with Australian, C-genome taxa in the ITS tree. In this case, neither of the two alternatives is especially well-supported, but the decay index is slightly higher for the conventional arrangement (i.e., a close relationship among African members of the genus; Fryxell 1979, 1992). The WSR tests suggest that the cpDNA and ITS data sets are strongly incongruent with respect to the position of the E-genome clade: the ITS data set strongly rejects ($p << 0.01$) monophyly for the (A + E + F)-genome clade, as found in the cpDNA tree, the cpDNA data reject the monophyly of the (E + C)-genome clade as found in the ITS phylogeny. The WSR tests, therefore, help to identify the E-genome clade as a contributor to total incongruence. Following exclusion of E-genome species from the data sets, I_{MF} changed slightly to 0.200 but the value of I_M increased to 0.576. Fewer nodes, however, were rejected by WSR tests (part 4 of Table 4; Fig. 9). It is surprising that removal of potentially problematic taxa had little impact on the magnitude of the incongruence indices, especially since removal of E-genome species resulted in less incongruence as judged by WSR tests (Table 4).

A less striking disagreement between the cpDNA and ITS data sets concerns the placement of subsection *Anomala*, which varies among the most-parsimonious cpDNA trees and hence is shown as part of a trichotomy in the strict consensus (Fig. 8A), but is strongly supported as a member of an African diploid clade (with A- and F-genome species) in the ITS trees (Fig. 8B). In this case, a monophyletic, African, (A + B + F)-genome clade, as exhibited by the ITS tree, is not rejected by the cpDNA data ($p = 0.15$). In addition, decay and character support values are high for this clade, whereas support for the alternative topology is less than half as strong. We therefore view the B-genome incongruence between the cpDNA and ITS data sets as more apparent than real, and we accept a robust African (A + B + F)-genome clade (Fig. 8C) as best reflecting our understanding of the phylogenetic position of subsection *Anomala*.

Within subgenus *Houzingenia*, the D-genome clade, topologies from analyses of separate data sets reveal similar groups of species, although in several cases the degree of resolution differs. For example, in both the cpDNA and ITS strict consensus trees, subsections *Caducibracteolata* (*G. armourianum* Kearney, *G. harknessii* Brandegee, *G. turneri* Fryxell), *Integrifolia* (*G. klotzschianum* Andersson and *G. davidsonii*), and *Houzingenia* [*G. thurberi* Tod. and *G. trilobum* (Moc & Seesé ex DC) Skvost.] are monophyletic. Perhaps more substantive are the incongruences observed among species groups, which are largely accounted for by the varying placements of subsection *Integrifolia*, subsection *Erioxylum* [*G. laxum* L. Ll. Phillips, *G. lobatum* Gentry, and *G. schwendimanii* Fryxell & S. D. Hoch (with *G. aridum* justifiably excluded)], and subsection *Austroamericana* in the two trees. In addition, WSR tests failed to reject alternatives found in the rival tree (nodes d1-d4 in part 3 of Table 4). We conclude, therefore, that we have insufficient evidence to regard the topologies as being statistically different, and therefore view the arrangement derived from the total evidence analysis (Fig. 8C) as the best available hypothesis of relationships. It should be noted, however, that within the D-genome clade of the total-evidence topology, support is weak for many interior branches, even though the topology is fully resolved.

A final example of phylogenetic incongruence in *Gossypium* concerns relationships among sections with subgenus *Sturtia* (C-genome). As in the previous example, support is weak for many of the alternative resolutions in the cpDNA and ITS-based phylogenies, raising the possibility that one or more of the observed incongruences is spurious or is due to insufficient character evolution. This is underscored by the failure of WSR tests to reject alternative topologies in most cases. The single prominent exception concerns the morphologically distinctive and geographically isolated section *Grandicalyx* (Fryxell 1979, 1992; Fryxell et al. 1992), which is strongly supported as a monophyletic group in the ITS tree (minimum of 10 synapomorphies and a decay index of 6), in accordance with expectations, but is paraphyletic in the cpDNA tree. In this case the ITS data are highly significantly incongruent with the cpDNA topology ($p << 0.01$).

Causes of phylogenetic incongruence. We have shown that once the known reticulate species have been pruned from the cpDNA and ITS data sets, substantial phylogenetic incongruence remains in the topologies produced. This incongruence is evident at all levels, ranging from the earliest divergence in the genus to more recent separations between species and species-groups. We explored the five most prominent examples of phylogenetic incongruence using comparative evaluation of support and decay in conjunction with statistical testing of constraint trees derived from the rival data sets. Among the five examples discussed, four represent cases wherein one resolution is strongly supported by one data set but only weakly supported or not supported by the rival data set: C-genome as basal; B-genome as a component of an African clade; positions of subsections within the D-genome clade; and the monophyly of section *Grandicalyx* within the C-genome clade. In each of these examples of "soft incongruence," the two alternative topologies, if considered in isolation, appear strikingly incongruent, but in the context of statistical tests and support measures, the incongruence appears to reflect insufficient information in one of the two data sets, and hence an unstable placement, and considerable character evolution and a robust resolution in the other. It seems prudent to posit this simple explanation as the cause of the conflict rather than to prematurely invoke one or more of the biological processes

capable of creating different historical records in two data sets, all the while recognizing that these are not ruled out.

These examples from *Gossypium* highlight a phenomenon that may be fairly common. It is widely appreciated that for any particular phylogenetic inference, there may be significant character support with one data set but little to no support using an alternative data set (e.g., Mason-Gamer and Kellogg 1996). This might be expected, for example, in many cases where two different data sets are generated using nucleotide sequences that differ greatly in substitution rates. Moreover, divergence events that are temporally closely spaced relative to the scale of accumulation of character differences for one or more data sets can lead to the same problem. Obtaining robust resolutions is often difficult for these cases involving short internodes, be they deep within the tree or relatively terminal, because the amount of character support is likely to be limited. Recovering identical topologies from additional independent data sets, especially those that can withstand statistical scrutiny, should not be expected in such cases. We suspect that this short internode phenomenon is a common cause of misleading phylogenetic inference as well as phylogenetic incongruence. In the present context, it may be evidenced by some of the incongruent resolutions obtained when two data sets were applied to both tribal-level (e.g., *Cienfuegosia* as basal in the tribe) and generic level (e.g., C-genome as basal in *Gossypium*) analyses. In these cases and others, the relevant internodes are supported by few characters and decay in trees only one to several steps longer than the most-parsimonious trees. This short internode phenomenon is also implicated by the molecular clock calculations (Table 5), which show (given the caveats and assumptions discussed above) that part of the difficulty in obtaining robust resolution among major clades in *Gossypium*, for example, stems from closely spaced divergences: the C-, D-, and (A + F)-genome clades are suggested to have diverged from each other between 10 and 12 million years ago.

The distinction between the two explanations offered so far, insufficient data and short internodes, is a fine one. Inasmuch as the short internode phenomenon often leads to limiting character support, it

represents a subset of the broader explanation of "insufficient data." In addition, though, it focuses attention on one of many possible biological phenomena that underlie phylogenetic incongruence. These include well-known processes at both the genomic and organismal levels. Examples of the former include gene recombination (Hein 1993), interlocus concerted evolution (Sang et al. 1995; Wendel et al. 1995b), and mistaken orthology assumptions (Doyle 1992; Sanderson and Doyle 1992), whereas examples of the latter include cytoplasmic and nuclear introgression (Rieseberg and Soltis 1991; Wendel et al. 1995a; Rieseberg et al. 1996), lineage sorting (Maddison 1995), and the short internode phenomenon, as described here. In *Gossypium*, a number of these processes have been described and have been suggested to be responsible for the incongruence exhibited by the nine "reticulate" taxa included in this study. Because several biological processes may lead to the same phylogenetic "phenotype," however (e.g., introgression vs. lineage sorting; Mason-Gamer et al. 1995; Wolfe and Elisens 1995), determining the actual cause of the incongruence will prove challenging in many cases.

A case in point is the E-genome clade in *Gossypium*, which, as described above, appears in a conventional position along with other African groups in the cpDNA-based tree but with Australian members of the genus in the ITS-based phylogeny. Support for each arrangement in the separate analyses is not convincing, but the conflict is statistically significant according to WSR tests. These results suggest, if only tentatively, that the incongruence reflects different underlying histories for the cpDNA and ITS data sets rather than insufficient or inadequate data. Additional insight may derive from the use of more data sets, but even if there emerged a majority view, the incongruence between the ITS sequence and cpDNA restriction site information would remain. Elucidating the cause might be problematic, and indeed, the possibility that the conflict is spurious cannot at present be excluded. Alternatively, we might meekly propose that the incongruence reflects peculiarities in the evolution of the ribosomal ITS sequences, which are known to be subject to a number of non-random molecular evolutionary processes that affect phylogenetic reconstruction, including concerted evolution (Baldwin et al. 1995), interlocus interactions

(Sang et al. 1995; Wendel et al. 1995a, 1995b), and possibly recombination with pseudogenes (Buckler and Holtsford 1996).

In summary, when faced with competing phylogenies, it seems rational to attempt to identify the components that are incongruent prior to combining data into a single "total evidence" analysis. Once these components have been identified, each may be independently evaluated with respect to the amount of conflict, in recognition of the fact that not all cases of incongruence are similar in magnitude or causation. This process (Fig. 10) may help to identify instances where conflict is more apparent than real and may also serve to focus attention on situations where the incongruence is sufficiently severe to warrant a search for its underlying biological causes. We suggest that this process will often yield insights into evolutionary history, and we have attempted to illustrate this with examples of competing phylogenetic inferences from the tribe Gossypieae and the genus *Gossypium*.

ACKNOWLEDGMENTS

We thank E. Kellogg for advice concerning application of WSR tests; J. La Duke for DNA samples of *Hibiscus costatus* and *Anotea flavida*; and the Institute for the Promotion of Teaching Science and Technology (Thailand) and the National Science Foundation (USA) for financial support. We are especially grateful to E. Kellogg and R. Mason-Gamer, whose work on congruence and incongruence in the Triticeae greatly influenced the present study, and their generous comments during initial review. We also would like to thank L. Clark, J. Davis, P. Fryxell, E. Kellogg, J. La Duke, R. Mason-Gamer, and R. Small for their comments on the manuscript.

LITERATURE CITED

ALEFELD, F. G. C. 1861. Ueber die Stellung der Gattung *Gossypium* und mehrer Anderer. Botanische Zeitung 19: 229-301.

BALDWIN, B. G., C. S. CAMPBELL, J. M. PORTER, M. J. SANDERSON, M. F. WOJCIECHOWSKI, and M. J. DONOGHUE. 1995. The ITS regions of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Annals of the Missouri Botanical Garden 82: 247-277.

BARRETT, M., M. J. DONOGHUE, and E. SOBER. 1991. Against consensus. Systematic Zoology 40: 486-493.

BENTHAM, G. and J. D. HOOKER. 1862-1867. *Genera plantarum.* vol. 1. Pp. 206-209. London: Reeve & Co.

BRUBAKER, C. L. and J. F. WENDEL. 1994. Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). American Journal of Botany 81: 1309-1326.

BUCKLER, E. S. and T. P. HOLTSFORD. 1996. *Zea* systematics: ribosomal ITS evidence. Molecular Biology and Evolution 13: 612-622.

BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, and P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. Systematic Biology 42: 384-397.

CLARK, L. G., W. P. ZHANG, and J. F. WENDEL. 1995. A phylogeny of the grass family (Poaceae) based on *ndhF* sequence data. Systematic Botany 20: 436-460.

DE QUEIROZ, A., M. J. DONOGHUE, and J. KIM. 1995. Separate versus combined analysis of phylogenetic evidence. Annual Review of Ecology and Systematics 26: 657-681.

DEGENER, O. 1934. *Kokia.* In *Flora Hawaiiensis; The new illustrated flora of the Hawaiian Islands.* Fam. 221., eds. O. Degener and L. D. Whitney. Honolulu.

DEJOODE, D. R. 1992. *Molecular insights into speciation in the genus Gossypium L. (Malvaceae).* M.S. thesis, Iowa State University, Ames.

—, and J. F. WENDEL. 1992. Genetic diversity and origin of the Hawaiian Islands cotton, *Gossypium tomentosum.* American Journal of Botany 79: 1311-1319.

DONOGHUE, M. J., R. G. OLMSTEAD, J. F. SMITH, and J. D. PALMER. 1992. Phylogenetic reationships of Dipsacales based on *rbcL* sequences. Annals of Missouri Botanical Garden 79: 333-345.

DOYLE, J. J. 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. Systematic Botany 17: 144-163.

DUMONT, A. 1887. Recherches sur l'anatomie comparée des Malvacées, Bombacacées, Tiliacées, Sterculiacées. Annales des Sciences Naturelles Seotiéne Serié 6: 129-246.

EDLIN, H. L. 1935. A critical revision of certain taxonomic groups of the Malvales part II. New Phytologist 34: 122-143.

ENDRIZZI, J. E., E. L. TURCOTTE, AND R. J. KOHEL. 1985. Genetics, cytology, and evolution of *Gossypium*. Advances in Genetics 23: 271-375.

FAITH, D. P. 1991. Cladistic permutation tests for monophyly and nonmonophyly. Systematic Zoology 40: 366-375.

FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, and C. BULT. 1995. Testing significance of incongruence. Cladistics 10: 315-319.

FOSBERG, F. R., and M. H. SACHÉT. 1966. *Lebronnecia*, gen. nov. (Malvaceae) des Iles Marquises. Adansonia n.s. 6:507-510.

FRYXELL, P. A. 1968. A redefinition of tribe Gossypieae. Botanical Gazette 129: 296-308.

—. 1969a. The genus *Hampea* (Malvaceae). Brittonia 21: 359-396.

—. 1969b. The genus *Cienfuegosia* Cav. (Malvaceae). Annals of the Missouri Botanical Garden 56: 179-250.

—. 1979. *The natural history of the cotton tribe*. College Station: Texas A & M University Press.

—. 1992. A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea 2: 108-165.

—, L. A. CRAVEN and J. MCD. STEWART. 1992. A revision of *Gossypium* sect. *Grandicalyx* (Malvaceae), including the description of six new species. Systematic Botany 17: 91-114.

GIBBONS, J. D. 1985. *Nonparametric statistical inference*, 2nd ed. New York: Marcel Dekker.

HARTER, H. L. and D. B. OWEN. 1970. *Selected tables in mathematical statistics*. Chicago: Markhan.

HEIN, J. 1993. A heuristic method to reconstruct the history of sequences subject to recombination. Journal of Molecular Evolution 36: 396-405.

HILLIS, D. M. 1987. Molecular versus morphological approaches to systematics. Annual Review of Ecology and Systematics 18: 23-42.

—, B. K. MABLE, and C. MORITZ. 1996. Applications of molecular systematics. Pp. 515-543 in *Molecular sytematics*, eds. D. M. Hillis, C. Moritz, and B. K. Mable, Sunderland, MA: Sinauer.

HUTCHINSON, J. B. 1943. A note on *Gossypium brevilanatum* Hochr. Tropical Agriculture 20: 4.

—. 1947. Notes on the classification and distribution of genera related to *Gossypium*. New Phytologist 46: 123-141.

KEARNEY, T. H. 1951. The American genera of Malvaceae. American Midland Naturalist 46: 93-131.

KELCHNER, S. A. and J. F. WENDEL. 1996. Hairpins create minute inversions in non-coding regions of chloroplast DNA. Current Genetics 30: 259-262.

KELLOGG, E. A., R. APPELS, and R. MASON-GAMER. 1996. When genes tell different stories: incongruent gene trees for diploid genera of Triticeae (Gramineae). Systematic Botany

KIM, K.-J., and R. K. JANSEN. 1994. Comparison of phylogenetic hypotheses among different data sets in dwarf dandelions (*Krigia*, Asteraceae): additional information from internal transcribed spacer sequences of nucear ribosomal DNA. Plant Systematics and Evolution 190: 157-185.

—. 1995. *ndhF* sequence evolution and the major clades in the sunflower family. Proceedings of the National Academy of Sciences U.S.A. 92: 10379-10383.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111-120.

KLUGE, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidea, Serpentes). Systematic Zoology 38: 7-25.

LA DUKE, J. C., and J. DOEBLEY. 1995. A chloroplast DNA based phylogeny of the Malvaceae. Systematic Botany 20: 259-271.

LEWTON, F. L. 1912. *Kokia*: a new genus of Hawaiian trees. Smithsonian Miscellaneous Collections 60 no. 5: 1-4 + 5 pls.

MACPHAIL, M. K., and E. M. TRUSWELL. 1989. Palynostratigraphy of the central west Murray Basin. Journal of Australian Geology and Geophysics 11: 301-331.

MADDISON, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. Systematic Zoology 40: 315-328.

MADDISON, W. P. 1995. Phylogenetic histories within and among species. Pp. 273-287 in *Experimental and molecular approaches to plant biosystematics*, eds. P. C. Hoch, A. G. Stevenson, and B. A. Schaal. St. Louis: Missouri Botanical Garden.

MASON-GAMER, R. J., and E. A. KELLOGG. 1996. Testing for phylogenetic conflict among molecular data sets in the Triticeae (Gramineae). Systematic Biology 45: 524-545.

—, K. E., HOLSINGER and R. K. JANSEN. 1995. Chloroplast DNA haplotype variation within and among populations of *Coreopsis grandiflora* (Asteraceae). Molecular Biology and Evolution 12: 371-381.

MICKEVICH, M. F., and J. S. FARRIS. 1981. The implications of congruence in *Menidia*. Systematic Zoology 30: 351-370.

MIYAMOTO, M. M. 1985. Consensus cladograms and general classfications. Cladistics 1: 186-189.

—, and W. M. FITCH. 1995. Testing species phylogenies and phylogenetic methods with congruence. Systematic Biology 44: 64-76.

MULLER, J. 1981. Fossil pollen records of extant angiosperms. Botanical Review 47: 1-142.

—. 1984. Significance of fossil pollen for angiosperm history. Annals of the Missouri Botanical Garden 71: 419-443.

OLMSTEAD, R. G., and P. A. REEVES. 1995. Evidence for polyphyly of the Scrophulariaceae based on chloroplast *rbcL* and *ndhF* sequences. Annals of the Missouri Botanical Garden 82: 176-193.

—, and J. A. SWEERE. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. Systematic Biology 43: 467-481.

—, J. A. SWEERE, and K. H. WOLFE. 1993. Ninety extra nucleotides in the *ndhF* gene of tobacco chloroplast DNA: a summary of revisions to the 1986 genome sequence. Plant Molecular Biology 22: 1191-1193

OMLAND, K. E. 1994. Character congruence between a molecular and a morphological phylogeny for dabbling ducks (*Anas*). Systematic Biology 43: 369-386.

PALMER, J. D. 1991. Plastid chromosome: structure and evolution. Cell Culture and Somatic Cell Genetics of Plants 7B: 5-53.

PATERSON, A. H., C. L. BRUBAKER, and J. F. WENDEL. 1993. A rapid method for extraction of cotton (*Gossyium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Molecular Biology Reporter 11: 122-127.

PERCIVAL, A. E. 1987. *The national collection of Gossypium germplasm.* College Station, Texas: United States Department of Agriculture Southern Crop Series Bulletin no. 321.

REEVES, R. G. 1936. Comparative anatomy of the seeds of cotton and other malvaceous plants II: Hibisceae. American Journal of Botany 23: 394-405.

RIESEBERG, L. H. and D. E. SOLTIS. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evolutionary Trends in Plants 5: 65-84.

—, J. WHITTON, and C. R. LINDER. 1996. Molecular marker incongruence in plant hybrid zones and phylogenetic trees. Acta Botanica Neerlandica 45: 243-262.

SANDERSON, M. J. and J. J. DOYLE. 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: Concerted evolution, homoplasy and confidence. Systematic Biology 41: 4-17.

SANG, T., D. J. CRAWFORD, and T. F. STUESSY. 1995. Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: Implications for biogeography and concerted evolution. Proceedings of the National Academy of Sciences (USA) 92: 6813-6817.

SCOTLAND, R. W., J. A. SWEERE, P. A. REEVES, and R. G. OLMSTEAD. 1995. Higher-level systematics of Acanthaceae determined by chloroplast DNA sequences. American Journal of Botany 82: 266-275.

SHINOZAKI,K., M. OHME, M. TANAKA, T. WAKASUGI, N. HAYASHIDA, T. MATSUBAYASHI, N. ZAITA, J. CHUNWONGSE, J. OBOKATA, K. YAMAGUCHI-SHINOZAKI, C. OHTO, K. TORAZAWA, B. Y. MENG, M. SUGITA, H. DENO, T. KAMOGASHIRA, K. YAMADA, J. KUSUDA, F. TAKAIWA, A. KATO, N. TOHDOH, H. SHIMADA, AND M. SUGIURA.

1986. The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. EMBO Journal 5: 2043-2049.

SIEGEL, S. 1956. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

STEPHENS, S. G. 1966. The potential for long range oceanic dispersal of cotton seeds. American Naturalist 100: 199-210.

SWOFFORD. 1991. When are phylogeny estimates from molecular and morphology data incongruent? Pp. 295-333 in *Phylogenetic analysis of DNA sequences*, eds. M. M. Miyamoto and J. Cracraft. New York: Oxford Univ. Press.

—, D. L. 1993. PAUP: Phylogenetic analysis using parsimony, version 3.1. Champaign: Illinois Natural History Survey.

TAJIMA, F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135: 599-607.

TAKEZAKI, N., A. RZHETSKY, and M. NEI. 1995. Phylogenetic test of the molecular clock and linearized trees. Molecular Biology and Evolution 12: 823-833.

TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37: 221-244.

WENDEL, J. F. 1989. New World tetraploid cottons contain Old World cytoplasm. Proceedings of the National Academy of Sciences U.S.A. 86: 4132-4136.

—, and V. A. ALBERT. 1992. Phylogenetics of the cotton genus (*Gossypium*): Character state weighted parsimony analysis of chloroplast DNA restriction site data and its systematic and biogeographic implication. Systematic Botany 17: 115-143.

—, A. SCHNABEL, and T. SEELANAN. 1995a. An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. Molecular Phylogenetics and Evolution 4: 298-313.

—, A. SCHNABEL, and T. SEELANAN. 1995b. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proceedings of the National Academy of Sciences U.S.A. 92: 280-284.

—, J. McD. Stewart and J. H. Rettig. 1991. Molecular evidence for homoploid reticulate evolution in Australian species of *Gossypium*. Evolution 45:
694-711.

Wolfe, A. D., and W. J. Elisens. 1995. Evidence of chloroplast capture and pollen-mediated gene flow in *Penstemon* sect. *Peltanthera* (Scrophulariaceae). Systematic Botany 20: 395-412.

TABLE 1. Plant materials used for obtaining ITS and *ndhF* sequences. The taxonomic treatment follows Fryxell (1979, 1992). ITS sequences were determined for all species listed, and *ndhF* sequences were generated for selected taxa, these denoted by underlined GenBank numbers. Accession names, where listed, are our own designations or are from the National Collection of *Gossypium* Germplasm (Percival 1987). Vouchers are deposited at ISC, with voucher abbreviations as follows: AAM = A. A. Mitchell; CI = I. Cowie; CLB = C. L. Brubaker; DDJ = Daniel DeJoode; EE = E. Edwards; JFW = J. F. Wendel; JIC = Juan Ismael Calzada; JMS = J. McD. Stewart; LAC = L. A. Craven; PAF = Paul A. Fryxell; TDC = T. D. Couch; TS = Tosak Seelanan. In few cases, *ndhF* and ITS sequences were obtained from the same genome DNAs, hence only one voucher specimen was given in such cases.

Species	**Accession**	**GenBank #**	**Voucher**
GOSSYPIEAE			
Gossypium L.			
Subgenus *Gossypium*			
Section *Gossypium*			
Subsection *Gossypium*			
Genomic designation: A			
G. arboreum L.	A_2-124	U55331	JFW & TDC 249
	A_2-74	U12712	JFW & TDC 305
G. herbaceum L.	A_1-73	U12713	JFW 539

TABLE 1 (continued).

Subsection *Anomala* Tod.			
Genomic designation: B			
G. anomalum Wawra ex Wawra & Peyr.	W4	U55332	JFW & TDC 312
		U56806	
G. capitis-viridis Mauer		U56807	JFW & TDC 609
		U56808	
Subsection *Pseudopambak* (Prokh.) Fryxell			
Genomic designation: E			
G. areysianum Deflers	E_3-1	U56810	TS 15
G. incanum (O. Schwartz) Hillc.	E_4-4	U56811	TS 7
G. somalense (Gürke) J. B. Hutch.	W43a	U56809	TS 14
G. stocksii Mast.		U55337	TS 13
		U56812	
Subsection *Longiloba* Fryxell			
Genomic designation: F			
G. longicalyx J. B. Hutch. & B. J. S. Lee	#191	U55338	TS 8
	F_1-1	U12722	
Section *Triphylla* (Prokh.) Fryxell			

TABLE 1 (continued).

Genomic designation: B			
G. triphyllum (Harv.) Hochr.	B_2-1	U12721	TS 16
Subgenus *Sturtia* (R. Br.) Tod.			
Genomic designation: C			
Section *Sturtia*			
G. robinsonii F. Mueller	AZ-50	U55334	TS 12
		U12710	
G. sturtianum J. H. Willis		U12720	JFW & TDC 216
Section *Hibiscoidea* Tod.			
G. australe F. Mueller	LC 9361	U56786	CLB & JFW 9361
G. bickii Prokh.	W56	U56787	JFW & TDC 557
G. nelsonii Fryxell	LC 9356	U56789	CLB & JFW 9356
Section *Grandicalyx* (Fryxell) Fryxell			
G. costulatum Tod.	PI 499772	U56790	PAF, LAC & JMS 3861
G. cunninghamii Tod.	PI 499774	U55333	PAF, LAC & JMS 4921
		U56788	
G. enthyle Fryxell, Craven & J. M. Stewart	LC 9224	U56791	LAC, JMS, & CLB 9224
G. exiguum Fryxell, Craven & J. M. Stewart	IC 4303	U56793	

TABLE 1 (continued).

	PI 499770	U56795	CI & JMS 4303
	PI 499786	U56798	JMS, JFW, & EE 9252
	PI 499793	U56792	PAF, LAC & JMS 4731
G. londonderriense Fryxell, Craven & J. M. Stewart	LC 9194	U56794	LAC, JMS, & JFW 9194
G. marchantii Fryxell, Craven & J. M. Stewart	LC 9170	U56796	JMS & CLB 9170
G. nobile Fryxell, Craven & J. M. Stewart	IC 4229	U56797	CI & JFW 4229
G. pilosum Fryxell	LC 9252	U56799	JMS & CLB 9252
G. populifolium (Benth.) F. Mueller ex Tod.	LC 9203	U56801	LAC, JMS, & CLB 9203
	PI 499808	U56800	PAF, LAC & JMS 4856
G. pulchellum (C. A. Gardner) Fryxell	PI 464858	U56802	JMS 20
G. rotundifolium Fryxell, Craven & J. M. Stewart	LC 9281	U56804	LAC, JMS, & AAM 9281
	PI 499789	U56803	JMS *s.n.*
G. sp. Nov.	LC 9191-2	U56805	LAC, JMS, & JFW 9191

Subgenus *Houzingenia* Fryxell

Genomic designation: D

Section *Houzingenia*

Subsection *Houzingenia*

TABLE 1 (continued).

G. thurberi Tod.	T#17	U12711	JFW & TDC 512
G. trilobum (Moç & Sessé ex DC.) Skovst.	D_8-1	U12723	JFW & TDC 507
Subsection *Integrifolia* (Tod.) Tod.			
G. davidsonii Kellogg	D_{3d}-32a	U12729	JFW & TDC 164
G. klotzschianum Andersson	D_{3k}-3	U12728	JFW & TDC 467
Subsection *Caducibracteolata* Mauer			
G. armourianum Kearney		U12725	TS 9
G. harknessii Brandegee		U12727	TS 10
G. turneri Fryxell	$D_{10\text{-}3}$	U55336	TS 11
		U12726	
Section *Erioxylum* (Rose & Standl.) Prokh.			
Subsection *Selera* (Ulbr.) Fryxell			
G. gossypioides (Ulbr.) Standl.	D_6-5	U12724	JFW & TDC 452
Subsection *Erioxylum* (Rose & Standl.) Fryxell			
G. aridum (Rose & Standl.) Skovst.	DRD108	U12732	DDJ & JIC 108
	DRD169	U12733	DDJ & JIC 169
G. laxum L. Ll. Phillips	DRD101	U12730	DDJ & JIC 101
G. lobatum Gentry	DRD161	U12731	

TABLE 1 (continued).

G. schwendimanii Fryxell & S. D. Koch	JMS	U12734	DDJ & JIC 150
Subsection *Austroamericana* Fryxell			
G. raimondii Ulbr.	# 436	U55335	JFW & TDC 127
	D_5-37	U12718	JFW & TDC 591
Subgenus *Karpas* Raf.			
Genomic designation: AD			
G. barbadense L.	K101	U55339	JFW & TDC 612
		U12715	
G. hirsutum L.	Palmeri	U55340	JFW & TDC 632
		U12719	
G. tomentosum Nuttall ex Seemann	WT936	U12717	JFW & TDC 621
G. mustelinum Miers ex Watt	W400	U12714	JFW & TDC 622
G. darwinii Watt	WB1215	U12716	JFW & TDC 620
Cienfuegosia Cav.			
Subgenus *Cienfuegosia*			
Section *Cienfuegosia*			
Cienfuegosia tripartita (H.B.K.) Gürke		U55324	
		U56777	

TABLE 1 (continued).

Section *Robusta* Fryxell		
Cienfuegosia intermedia Fryxell	U56776	
Gossypioides Skovst. ex J. B. Hutch.		
Gossypioides kirkii (Mast.) J. B. Hutch.	U55329	TS 3
	U56783	
Hampea Schltdl.		
Section *Trianchonia* Fryxell		
Hampea appendiculata (Donn.Sm.) Standl.	U55327	Pohl 15736A
	U56781	TS 4
Kokia Lewton		
Kokia drynarioides (Seemann) Lewton	U55330	TS 6
	U56784	
Kokia kauaiensis (Rock) O. Deg. & Duvel	U56785	
Lebronnecia Fosberg		
Lebronnecia kokioides Fosberg	U55325	JFW & TDC 624
	U56778	
Thespesia Sol.		
Section *Thespesia*		

TABLE 1 (continued).

Thespesia populnea (L.) Sol. ex Correa		U55328 U56782	TS 1
Section *Lampas* (Ulbr.) Borss. Waalk.			
Thespesia lampas (Cav.) Dalzell ex Dalzell & Gibson			U56779 TS 2
Thespesia thespesioides (R. Br. ex Benth.) Fryxell		U55326 U56780	
MALVAVISCEAE			
Anotea			
Anotea flavida Ulbr.		U55322 U56775	Fryxell 785630
HIBISCEAE			
Hibiscus			
Hibiscus costatus A. Rich.	LC 9166	U55323 U56774	

TABLE 2. Synonymous (K_S) and nonsynonymous (K_A) substitution rates (per 100 sites) and percent sequence divergence for *ndhF* and *rbcL* genes in *Gossypium* and *Thespesia*. Percent sequence divergence estimates are corrected for multiple substitutions using the Kimura 2-parameter method (= K2P). *Gossypium* "D" = *Gossypium raimondii* and *Gossypium turneri*. For each comparison, the three values listed represent K_S (top), K_A (middle), and K2P (bottom). Estimates for *ndhF* involving *Gossypium* "D" are averages of values obtained for comparisons with *G. raimondii* and with *G. turneri*. TP = *Thespesia populnea*, GR = *Gossypium robinsonii*, GA = *Gossypium anomalum*, GD = *Gossypium davidsonii*, GH = *Gossypium hirsutum*, G "D" = *Gossypium raimondii* and *Gossypium turneri*.

ndhF

Taxon	TP	GR	GA	G "D"	GH
TP		3.28	2.82	3.28	2.94
		0.89	0.83	0.95	0.99
		1.43	1.28	1.48	1.43
GR			0.21	1.62	1.84
			0.57	0.86	0.73
			0.49	1.03	0.98
GA				1.19	1.40
				0.74	0.67
				0.83	0.83
G "D"				0.64	1.73
				0.32	0.96
				0.39	1.13

rbcL

Taxon	TP	GR	GA	GD	GH
TP		3.35	1.81	1.81	2.74
		0.66	0.66	0.76	1.23
		1.30	0.94	1.01	1.59
GR			1.50	1.50	1.81
			0.19	0.28	0.76
			0.50	0.57	1.01
GA				0.00	0.90
				0.28	0.95
				0.21	0.94
GD					0.90
					0.66
					0.72

TABLE 3. Character incongruence between data sets in the tribe Gossypieae and in *Gossypium*, for cases where known or suspected reticulate taxa are included or excluded from the analysis. Data sets used are sequences for the plastid gene *ndhF* and the nuclear rDNA internal transcribed spacer (ITS), and the cpDNA restriction sites from Wendel and Albert (1992). Abbreviations for tree statistics and calculations are as follows: L = length of most-parsimonious tree(s); R = minimum number of synapomorphies in most-parsimonious tree(s); S = number of extra steps; #Tree = number of most-parsimonious trees; CI = consistency index; RI = retention index; L^H = length of shortest tree(s) when characters from one data set are optimized on a constraint tree from the other data set; E = number of extra steps (= L^H-R); I_{MF} = Mickevich-Farris index; I_M = Miyamoto index. Values in parentheses were obtained when *G. anomalum* was excluded from data sets along with the reticulate taxa.

Tribe Gossypieae

	All taxa included			Reticulate taxa excluded					
	ndhF	ITS	Combined	*ndhF*		ITS		Combined	
L	221	678	917	208	(205)	649	(624)	863	(831)
R	202	488	690	192	(189)	477	(464)	669	(653)
S	19	190	227	17	(16)	172	(160)	194	(178)
#Tree	15	4	5	15	(15)	4	(4)	5	(4)
CI	0.914	0.720	0.752	0.919	(0.922)	0.735	(0.744)	0.775	(0.786)
RI	0.898	0.603	0.659	0.886	(0.885)	0.598	(0.602)	0.664	(0.671)
L^H	243	718	—	214	(207)	664	(628)	—	—
E	41	230	—	22	(18)	187	(164)	—	—

TABLE 3 (continued).

I_{MF}	0.079	0.026 (0.011)
I_M	0.229	0.096 (0.033)

Genus *Gossypium*

	All taxa included			Reticulate taxa excluded		
	cpDNA	ITS	Combined	cpDNA	ITS	Combined
L	228	319	632	199	266	485
R	202	236	438	175	208	383
S	26	83	194	24	58	102
#Tree	4	90	10	8	12	1
CI	0.886	0.740	0.693	0.879	0.782	0.790
RI	0.964	0.848	0.848	0.955	0.854	0.891
L^H	372	479	—	242	318	—
E	170	243	—	67	110	—
I_{MF}		0.438			0.196	
I_M		0.736			0.537	

TABLE 4. Wilcoxon signed rank tests for incongruence between data sets. Gain, loss, and net refer to the number of extra steps, the number of fewer steps, and the total between a constraint tree and an unconstrained tree, respectively (see text for details). Probability values (*P*) are shown, except when only 1 character was different between trees or when there was no length difference, in which cases the WSR test cannot be applied.

Data set/constraint tree	Gain	Loss	Net	*p*
1) Gossypieae (see Fig. 6 for constraint trees and nodes)				
A) All taxa included				
ITS				
ndhF strict consensus tree	46	7	39	$p \ll 0.01$
(i) monophyly of *Lebronnecia-Hampea-Thespesia*	14	12	2	$p \gg 0.37$
(iv) monophyly of A-AD-F genomes	22	1	21	$p \ll 0.01$
(v) monophyly of B-C genomes	18	5	13	$p < 0.01$
combined tree	9	5	4	$p \gg 0.19$
ndhF				
ITS strict consensus tree	23	1	22	$p \ll 0.01$
(vi) *(Thespesia populnea, ((Gossypioides, Kokia), Gossypium))*	2	0	2	$p = 0.25$
(ix) monophyly of A-F-B genomes	11	0	11	$p \ll 0.01$
(x) monophyly of C-E-D-AD genomes	10	0	10	$p \ll 0.01$
(xi) monophyly of C-E genomes	3	1	2	$p = 0.25$
(xii) monophyly of D-AD genomes	9	0	9	$p \ll 0.01$
combined tree	12	0	12	$p \ll 0.01$

TABLE 4 (continued).

B) Reticulate taxa in *Gossypium* removed				
ITS				
(v-a) monophyly of B-C genomes	23	10	13	$p \approx 0.02$
ndhF				
(ix-a) monophyly of A-F-B genomes	3	0	3	$p \approx 0.13$
(x-a) monophyly of C-E-D genomes	3	1	2	$p = 0.25$
combined tree	4	1	3	$p \approx 0.15$
C) Reticulate taxa and *G. anomalum* in *Gossypium* removed				
ITS				
ndhF strict consensus tree	10	8	2	$p \approx 0.35$
ndhF				
ITS strict consensus tree	4	2	2	$p \approx 0.28$
2) *Gossypium* — all taxa (see Fig. 7 for constraint trees and nodes)				
ITS				
cpDNA strict consensus tree	118	1	117	$p < 0.01$
(i) monophyly of A-AD-E-F-B-D genomes	8	6	2	$p \approx 0.34$
(ii) monophyly of A-AD-E-F genomes	28	4	24	$p < 0.01$
(iii) monophyly of D genome	21	1	20	$p < 0.01$
(iii) B-genome sister to D-genome	24	2	22	$p < 0.01$
(v) first split of C genome	8	4	4	$p \approx 0.17$
combined tree	39	2	37	$p < 0.01$
cpDNA				
ITS strict consensus tree	137	0	137	$p < 0.01$

TABLE 4 (continued).

(vi) monophyly of A-AD-B-F-D genomes	36	0	36	$p \ll 0.01$
(vii) monophyly of C-E-D-AD genomes	41	0	41	$p \ll 0.01$
(viii) monophyly of D-AD genomes	35	0	35	$p \ll 0.01$
(ix) monophyly of C-E genomes	11	3	8	$p \gg 0.03$
(x) first split of ABF genomes	41	0	41	$p \ll 0.01$
combined tree	8	0	8	$p < 0.01$
3) *Gossypium* without reticulate taxa (see Fig. 8 for constraint trees and nodes)				
ITS				
cpDNA strict consensus tree	46	3	43	$p \ll 0.01$
(i) monophyly of AEFBD genomes	7	5	2	$p \gg 0.32$
(ii) monophyly of AEF genomes	17	2	15	$p \ll 0.01$
(iii) B-genome sister to D-genome	15	3	12	$p \ll 0.01$
(v) first split of C genome	7	3	4	$p \gg 0.15$
(c1) paraphyly of Section *Grandicalyx*	19	0	19	$p \ll 0.01$
(c4) Section *Sturtia* as a basal clade in C genome	2	0	2	$p = 0.25$
(c5) Section *Hibiscoidea* sister to Section *Grandicalyx*	2	0	2	$p = 0.25$
(d1) Subsection *Integrifolia* as a basal clade in D genome	3	0	3	$p \gg 0.13$
(d2) Subsection *Caducibracteolata* as a sister clade to the rest of D genome	3	0	3	$p \gg 0.13$
combined tree	16	4	12	$p < 0.01$
cpDNA				
ITS strict consensus tree	40	2	38	$p \ll 0.01$
(vi) monophyly of ABF genomes	7	3	4	$p \gg 0.15$
(vii) monophyly of CED genomes	12	3	9	$p \gg 0.02$

TABLE 4 (continued).

(x) first split of ABF genomes	12	3	9	$p \approx 0.02$
(c2) Section *Sturtia* sister to Section *Hibiscoidea*	9	2	7	$p \approx 0.03$
(c3) monophyly of Section *Grandicalyx*	1	0	1	—
(c4) Section *Sturtia* as a basal clade in C genome	—	—	—	—
(c5) Section *Hibiscoidea* sister to Section *Grandicalyx*	1	0	1	—
(d3) Subsection *Erioxylum* a basal clade in D genome	4	0	4	$p \approx 0.06$
(d4) Subsection *Caducibracteolata* sister to the rest of D genome	4	0	4	$p \approx 0.06$
combined tree	13	5	8	$p \approx 0.06$

4) *Gossypium* without reticulate taxa and E-genome (see Fig. 9 for constraint trees and nodes)

ITS

cpDNA strict consensus tree	30	2	28	$p < 0.01$
(i) AFBD monophyly	2	1	1	$p = 0.38$
(ii) AF monophyly	2	1	1	$p = 0.38$
(iii) AF/B/D trichotomy	5	3	2	$p \approx 0.30$
(iv) first split of C-genome	3	2	1	$p \approx 0.41$
(c1) paraphyly of section *Grandicalyx*	19	0	19	$p < 0.01$
(d1) Section *Integrifolia* as a basal clade in D-genome	3	0	3	$p \approx 0.13$
(d2) Section Caducibracteolata as a sister clade to the rest of D-genome	3	0	3	$p \approx 0.13$
(c5) Section *Sturtia* as a basal clade in C genome	2	0	2	$p = 0.25$
(c6) Section *Hibiscoidea* sister to Section *Grandicalyx*	2	0	2	$p = 0.25$
combined tree	12	3	9	$p \approx 0.02$

CpDNA

ITS strict consensus tree	36	2	34	$p < 0.01$

TABLE 4 (continued).

(v) monophyly of CD genomes	7	2	5	$p = 0.13$
(vi) first split of ABF genomes	7	2	5	$p \approx 0.08$
(c2) Section *Grandicalyx* as a basal clade in C-genome	9	2	7	$p \approx 0.03$
(c3) Section *Sturtia* sister to Section *Hibiscoidea*	9	2	7	$p \approx 0.03$
(c4) monophyly of Section *Grandicalyx*	1	0	1	—
(d3) Subsection *Erioxylum* a basal clade in D genome	4	0	4	$p \approx 0.06$
(d4) Subsection *Caducibracteolata* sister to the rest of D genome	4	0	4	$p \approx 0.06$
(c5) Section *Sturtia* as a basal clade in C genome	2	2	0	$p \approx 0.06$
(c6) Section *Hibiscoidea* sister to Section *Grandicalyx*	1	0	1	—
combined tree	6	2	4	$p = 0.13$

TABLE 5. Estimates of time since divergence (above diagonal, in mybp) for members of tribe Gossypieae based on Kimura (1980) two-parameter estimates of percent sequence divergence in *ndhF* gene (below diagonal). Taxon abbreviations are: CT = *Cienfuegosia tripartita*; LK = *Lebronnecia kokioides*; TT = *Thespesia thespesioides*; HA = *Hampea appendiculata*; TP = *Thespesia populnea*; GK = *Gossypioides kirkii*; KD = *Kokia drynarioides*; Ga = *Gossypium arboreum*; Gb = *Gossypium anomalum*; Gc = *Gossypium cunninghamii* & *Gossypium robinsonii*; Gd = *Gossypium raimondii* & *Gossypium turneri*; Ge = *Gossypium stocksii*; Gf = *Gossypium longicalyx*; Gad = *Gossypium barbadense* & *Gossypium hirsutum*. Lowercase letters in *Gossypium* refer to genome designation. When two species are listed above, the estimate shown is an average for the listed species.

	CT	LK	TT	HA	TP	GK	KD	Ga	Gb	Gc	Gd	Ge	Gf	Gad
CT	—	17	16	19	16	17	18	21	19	22	21	19	19	21
LK	1.73	—	6	10	9	12	13	16	15	18	17	14	14	16
TT	1.63	0.59	—	8	7	10	11	15	13	16	15	13	12	14
HA	1.88	1.03	0.79	—	10	14	15	18	16	19	19	15	16	18
TP	1.63	0.88	0.69	0.98	—	10	11	15	13	15	15	13	12	14
GK	1.68	1.23	1.03	1.38	1.03	—	3	13	11	14	13	11	10	12
KD	1.78	1.33	1.13	1.48	1.13	0.29	—	14	12	15	14	12	11	13
Ga	2.13	1.58	1.48	1.78	1.48	1.28	1.38	—	9	12	11	7	4	2
Gb	1.93	1.48	1.28	1.58	1.28	1.08	1.18	0.88	—	7	8	6	6	8
Gc	2.21	1.77	1.58	1.92	1.53	1.38	1.48	1.18	0.66	0.68/7	12	9	9	11
Gd	2.13	1.72	1.48	1.87	1.48	1.28	1.43	1.13	0.83	0.68	0.39/4	8	9	11
Ge	1.93	1.38	1.28	1.48	1.28	1.08	1.18	0.69	0.59	0.89	0.83	—	5	7
Gf	1.88	1.43	1.23	1.57	1.23	1.03	1.13	0.44	0.64	0.93	0.93	0.54	—	4
Gad	2.06	1.60	1.41	1.75	1.41	1.21	1.31	0.22	0.81	1.11	1.11	0.72	0.37	0.05/0.5

FIGURE LEGEND

FIG. 1. Distribution of variation by position for the plastid gene *ndhF* in the Gossypieae (outgroups excluded). Shown are the number of variable sites (dark bars) and potentially phylogenetically informative sites (light bars) in non-overlapping 100-bp intervals (the final interval spans 82 bp). Coordinates are from Shinozaki et al. (1986) as corrected by Olmstead and Sweere (1993) by the insertion of 90 bp that was missing from the original sequence (between 112,469 and 112,470, at asterisk). Lower case letters refer to gaps shown in Fig. 2.

FIG. 2. A portion of the matrix of aligned *ndhF* sequences illustrating the gaps detected in the 3' half of the gene. Lower case letters "a" through "e" identify the indels, which are mapped in Fig. 1 according to the coordinates shown (from Shinozaki et al. 1986).

FIG. 3. Topologies discovered in phylogenetic analyses of *ndhF* sequences in the Gossypieae. Trees are rooted with sequences from *Hibiscus costatus* and *Anotea flavida*. Top: parsimony analysis. Thin lines indicate positions held by *Cienfuegosia tripartita* (**A**, **B**, or **C**) and *Thespesia thespesioides* (**i** or **ii**) in alternative most-parsimonious trees and in analyses that differed in the treatment of gaps (see text). *Gossypium* species are shown as a solid triangle, which is expanded in the inset (G1-G6) to illustrate six alternative topologies recovered under various gap treatments (A - F and AD indicate genome designation). Numbers above branches indicate the minimum amount of character support in all most-parsimonious trees; two numbers are shown if the branch may be partitioned by the placement of *Cienfuegosia tripartita* or *Thespesia thespesioides*, and circled numbers indicate support when *Thespesia thespesioides* is at **i** and *Cienfuegosia tripartita* is at **A**. The "1" in the square box indicates the single character supporting the branch segment leading to *Lebronnecia kokioides* and *Thespesia thespesioides* when *Thespesia thespesioides* is at **i**. Bottom: neighbor-joining tree constructed from a matrix of Kimura 2P-corrected distances. Branch lengths are proportional to distances, and genome designation (A - F and AD) for each taxon are shown in parentheses after species.

FIG. 4. A portion of the matrix of aligned ITS sequences. Several large gaps are apparent near the start of ITS1 and ITS2, which have complete aligned lengths of 321 and 239 bp, respectively. *Gossypium australe* is arbitrarily shown as a representative of the 45 sequences from *Gossypium*.

FIG. 5. Tree topologies discovered in phylogenetic analyses of ITS sequences. Thin lines indicate possible placements of *Cienfuegosia* spp. (**a** or **b**) and *Gossypium* E-genome species (**i** through **iv**). *Cienfuegosia* spp. were at **a** and at **b** in 3,564 and 648 most-parsimonious trees, respectively. In the former, *Gossypium* E-genome species were located at any one of the positions **i** through **iv**, whereas if *Cienfuegosia* spp. occupied position **b**, *Gossypium* E-genome species were located at **i** or **iii** or **iv**. Numbers below branches indicate decay indices and those above branches show minimum character support. Support for the branch leading to **b** is the same as for branch **a**, and support for branches **ii**, **iii**, and **iv** is the same as for branch **i**. Triangles denote branches containing more than one species. The shaded region of the topology shows relationships among genomic groups within *Gossypium*.

FIG. 6. Strict consensus trees from analyses of *ndhF*, ITS, and combined data sets (A, B, and C, respectively). A. Strict consensus of the 15 most-parsimonious trees found in analyses of *ndhF* sequences, with gaps coded as missing data (uninformative characters excluded; length = 113; CI = 0.83; RI = 0.89). B. Strict consensus of the four most-parsimonious trees obtained from analyses of ITS sequences (length = 487; CI = 0.61; RI = 0.60). C. Strict consensus of four most-parsimonious trees resulting from analyses of a combined data matrix consisting of *ndhF* and ITS sequences (length = 618; CI = 0.63; RI = 0.65). *Hibiscus costatus* and *Anotea flavida* were used to root all trees. Upper case letters to the right of taxon names refer to genomic designations in *Gossypium*. Numbers below branches indicate decay indices and those above branches show minimum character support. Roman numerals designate nodes used for Wilcoxon Signed Rank tests of incongruence (see text).

FIG. 7. Strict consensus trees from analyses of cpDNA, ITS, and combined data sets (A, B, and C, respectively) when all taxa in Wendel and Albert (1992) are included. A. Strict consensus of four shortest trees found in analyses of cpDNA restriction site variation (uninformative characters

excluded; length = 161; CI = 0.84; RI = 0.81). B. Strict consensus of 90 most-parsimonious trees obtained from analyses of ITS sequences, using the same taxa as in A (length = 222; CI = 0.63; RI = 0.53). C. Strict consensus of the 10 shortest trees found in analyses of a combined matrix consisting of cpDNA restriction site and ITS data (length = 468; CI = 0.59; RI = 0.85). *Thespesia populnea* was used to root all trees. Upper case letters to the right of each cladogram refer to genomic designations. Numbers below branches indicate decay indices. Roman numerals designate nodes used for Wilcoxon Signed Rank tests of incongruence (see text).

FIG. 8. Strict consensus trees from analyses of cpDNA, ITS, and combined data sets (A, B, and C, respectively) when known or suspected reticulate taxa are excluded from the data matrix. A. Strict consensus of eight most-parsimonious trees found in analyses of cpDNA restriction site variation (uninformative characters excluded; length = 137; CI = 0.83; RI = 0.96). B. Strict consensus of 12 shortest trees obtained from analyses of ITS sequences (length = 179; CI = 0.68; RI = 0.85). C. The most-parsimonious trees recovered in an analysis of a combined matrix consisting of cpDNA restriction site and ITS sequence data (length = 342; CI = 0.70; RI = 0.89). *Thespesia populnea* was used to root all trees. Upper case letters to the right of each cladogram refer to genomic designations. Numbers below branches indicate decay indices and those above branches show minimum character support. Roman numerals and letters with numbers designate nodes used for Wilcoxon Signed Rank tests of incongruence (see text).

FIG. 9. Strict consensus trees from analyses of cpDNA, ITS, and combined data sets (A, B, and C, respectively) when reticulate taxa and E-genome species are excluded from the data matrix. A. Strict consensus of six most-parsimonious trees found in analyses of cpDNA restriction site variation (uninformative characters excluded; length = 116; CI = 0.85; RI = 0.97). B. Strict consensus of two shortest trees obtained from analyses of ITS sequences (length = 142; CI = 0.73; RI = 0.89). C. Strict consensus of the most-parsimonious tree recovered in an analysis of a combined matrix consisting of cpDNA restriction site and ITS sequence data (length = 271; CI = 0.75; RI = 0.92). *Thespesia populnea* was used to root all trees. Upper case letters to the right of each cladogram refer to genomic designations.

Numbers below branches indicate decay indices and those above branches show minimum character support. Roman numerals and letters with numbers designate nodes used for Wilcoxon Signed Rank tests of incongruence (see text).

FIG. 10. Suggested procedure for evaluating competing trees.

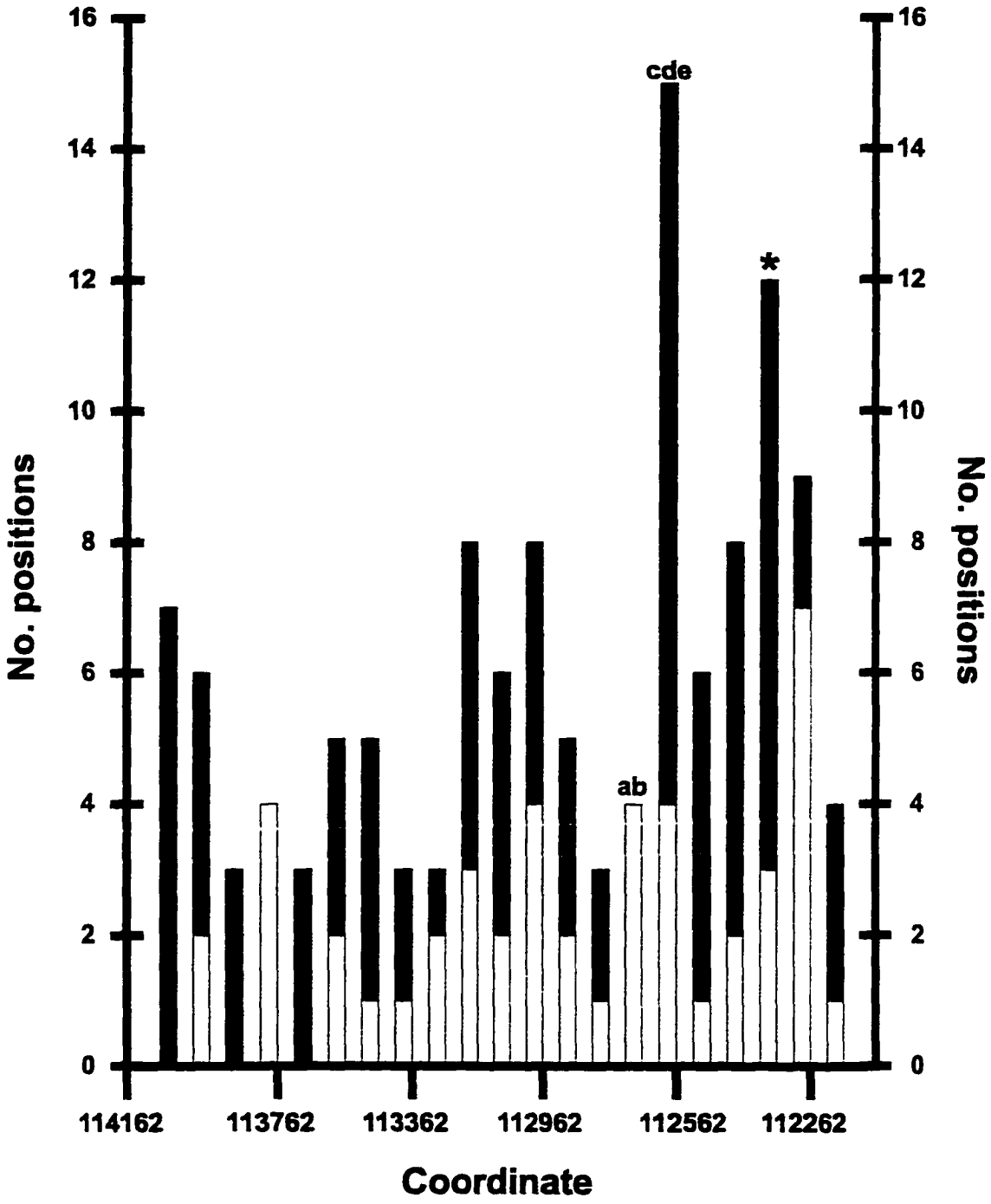

16
14
12
10
8
6
4
2
0
No. positions
cde
ab
*
114162
113762
113362
112962
112562
112262
Coordinate

Fig. 1

Coordinate 112,754 - 112,683:

Coordinate 112,682 - 112,611:

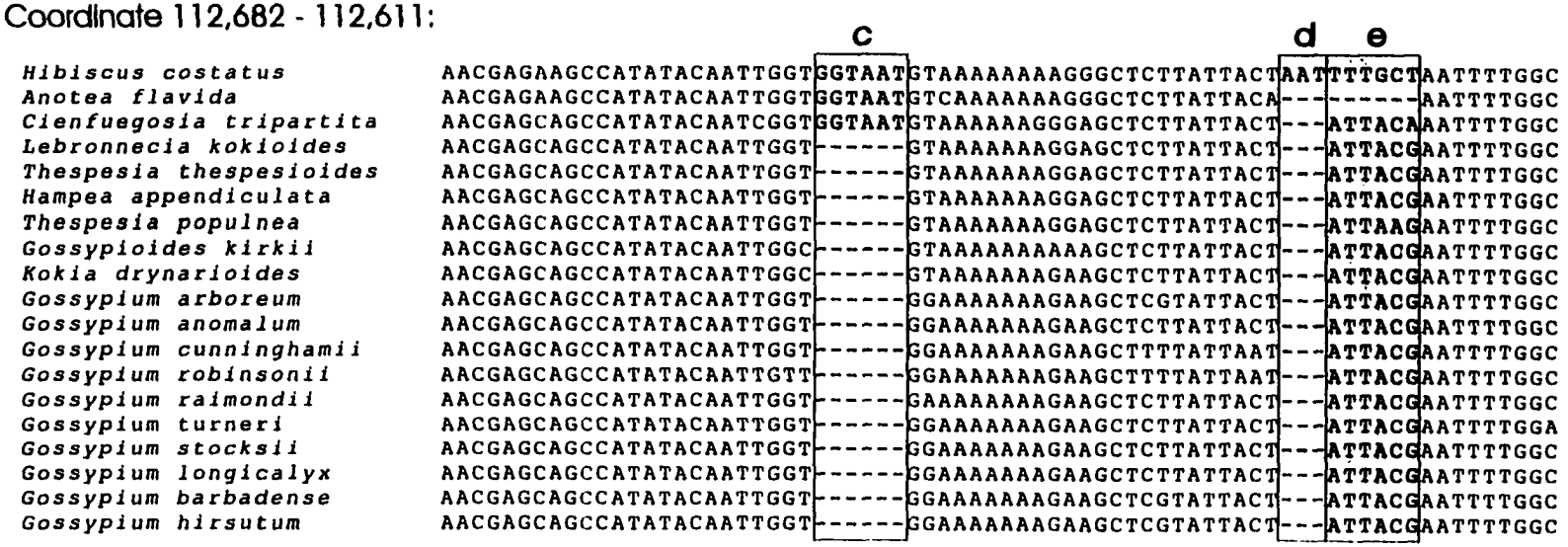
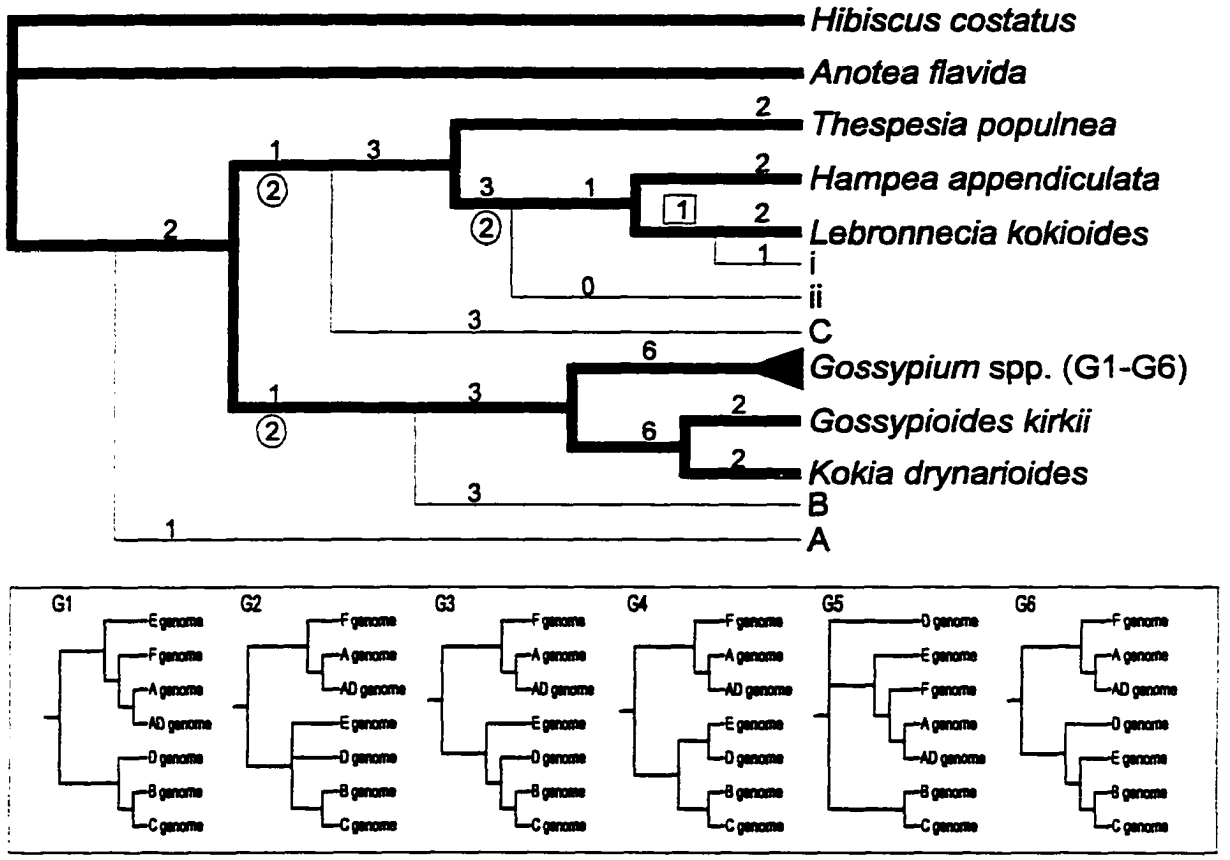


Fig. 2


Hibiscus costatus
Anotea flavida
Thespesia populnea
Hampea appendiculata
Lebronnecia kokioides
i
ii
C
Gossypium spp. (G1-G6)
Gossypioides kirkii
Kokia drynarioides
B
A
G1
G2
G3
G4
G5
G6

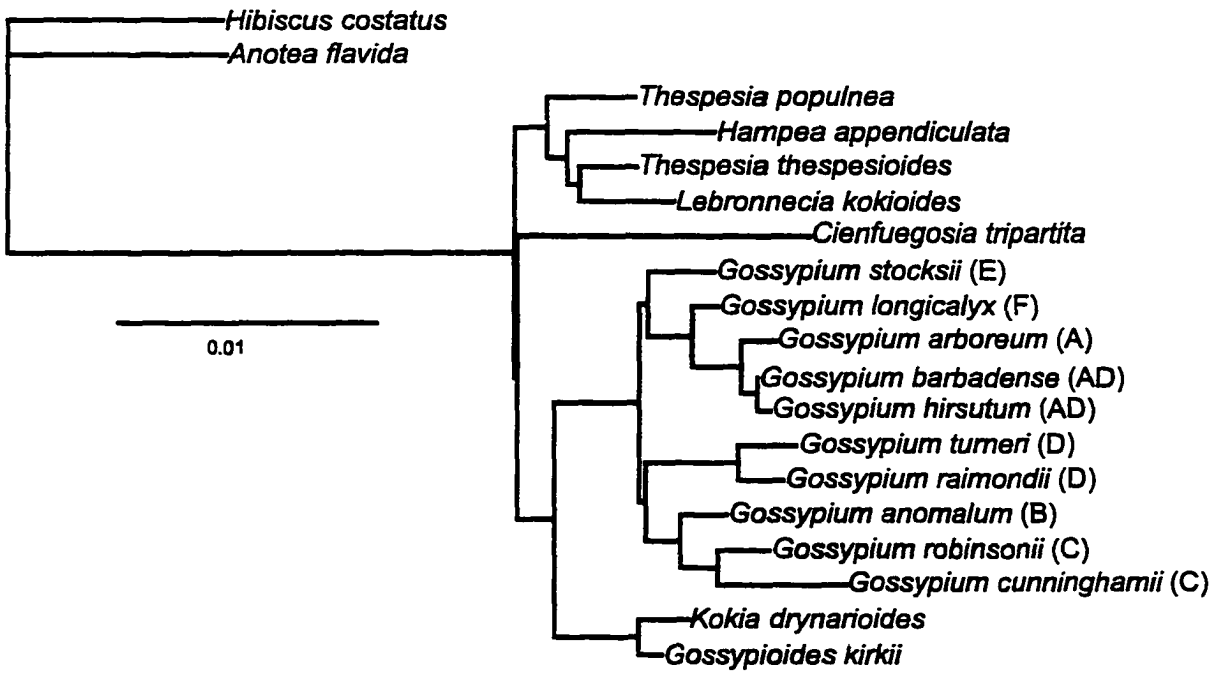
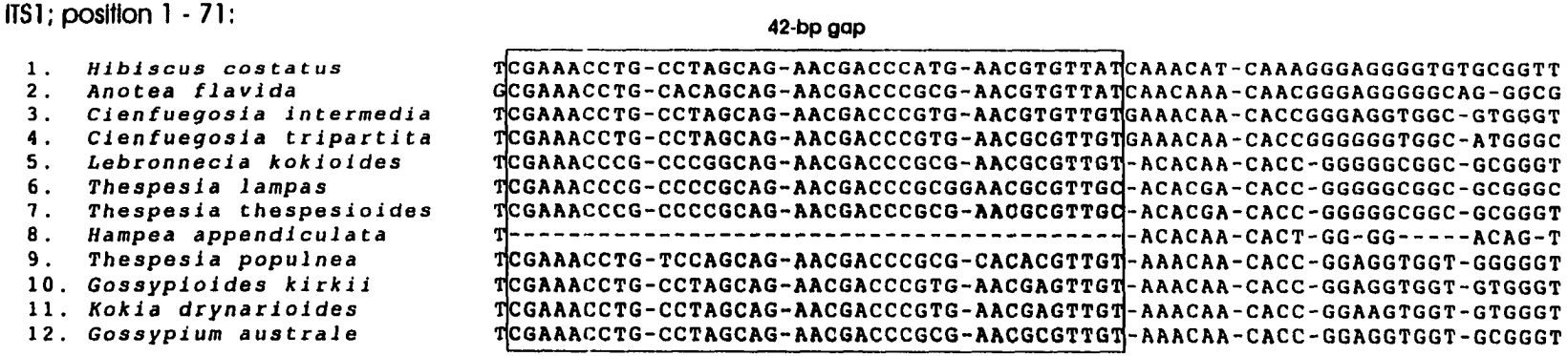

Hibiscus costatus
Anotea flavida
Thespesia populnea
Hampea appendiculata
Thespesia thespesioides
Lebronnecia kokioides
Cienfuegosia tripartita
Gossypium stocksii (E)
Gossypium longicalyx (F)
Gossypium arboreum (A)
Gossypium barbadense (AD)
Gossypium hirsutum (AD)
Gossypium turneri (D)
Gossypium raimondii (D)
Gossypium anomalum (B)
Gossypium robinsonii (C)
Gossypium cunninghamii (C)
Kokia drynarioides
Gossypioides kirkii
0.01

Fig. 3

ITS1; position 1 - 71:



ITS1; position 72 - 169:

ITS2; position 508 - 530:

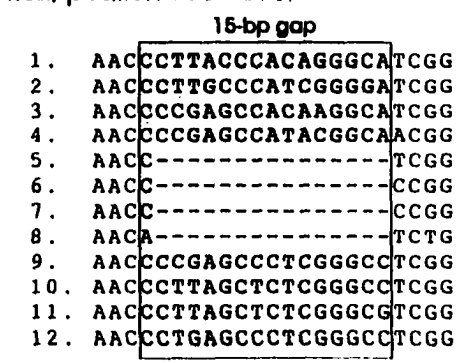


Fig. 4

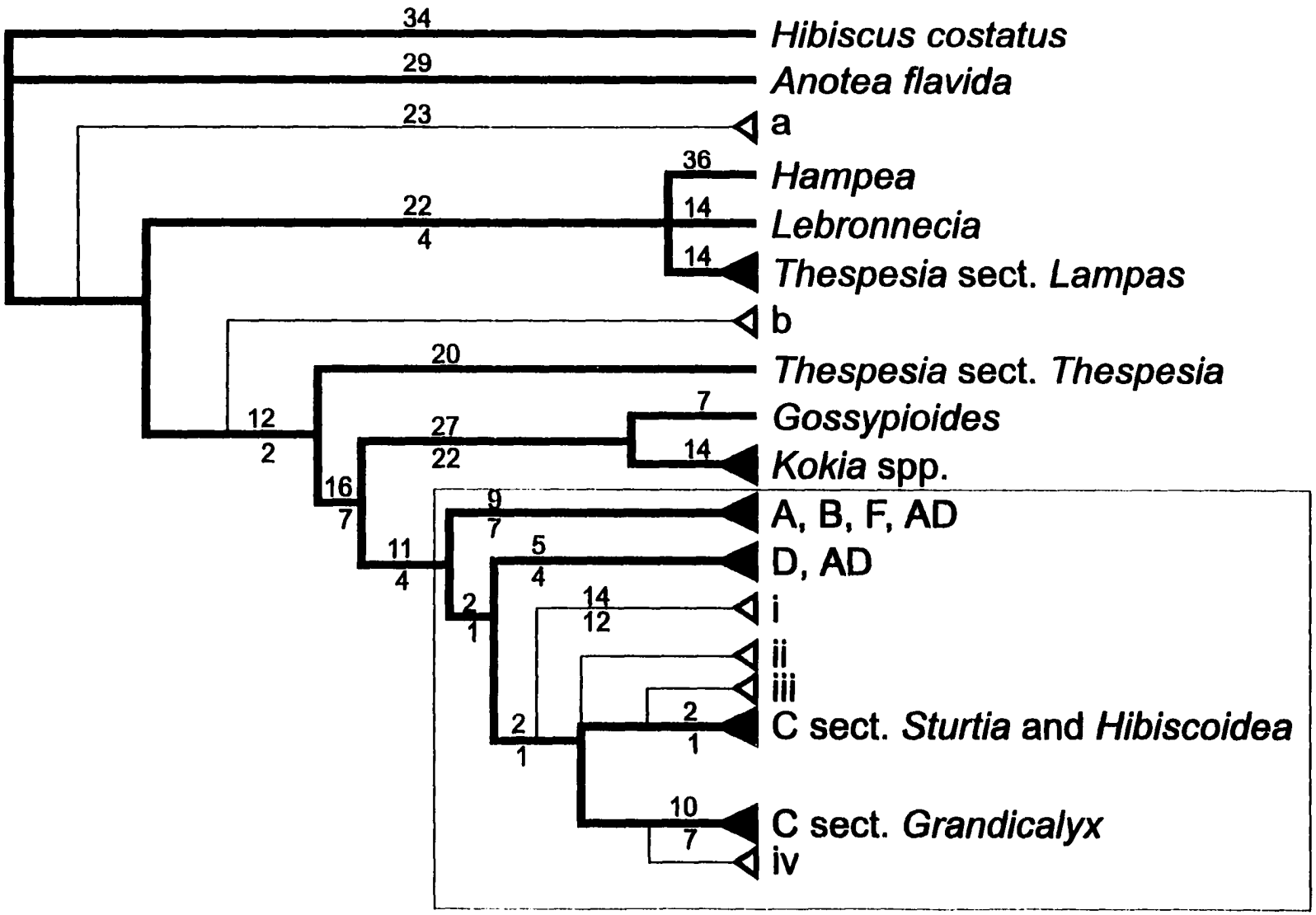

34
Hibiscus costatus
29
Anotea flavida
23
a
36
Hampea
22
14
Lebronnecia
4
14
Thespesia sect. Lampas
b
20
Thespesia sect. Thespesia
7
Gossypioides
12
27
14
Kokia spp.
2
22
16
9
A, B, F, AD
7
7
11
5
D, AD
4
4
2
14
i
1
12
ii
iii
2
2
C sect. Sturtia and Hibiscoidea
1
1
10
C sect. Grandicalyx
7
iv

Fig. 5

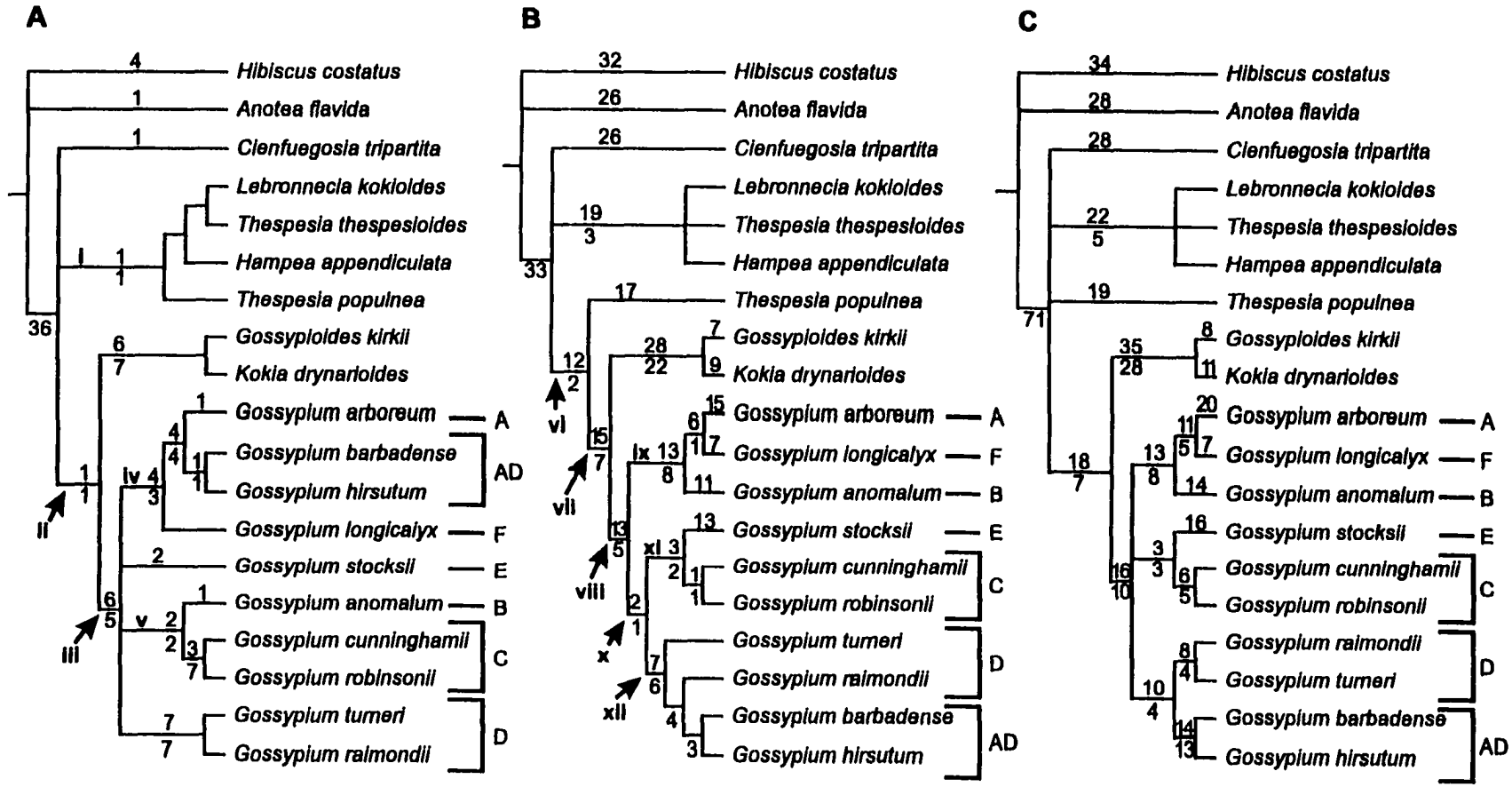

A
B
C
Hibiscus costatus
Anotea flavida
Clenfuegosia tripartita
Lebronnecia kokioides
Thespesia thespesioides
Hampea appendiculata
Thespesia populnea
Gossypioides kirkii
Kokia drynarioides
Gossypium arboreum
Gossypium barbadense
Gossypium hirsutum
Gossypium longicalyx
Gossypium stocksii
Gossypium anomalum
Gossypium cunninghamii
Gossypium robinsonii
Gossypium turneri
Gossypium raimondii
A
AD
F
E
B
C
D
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII

Fig. 6

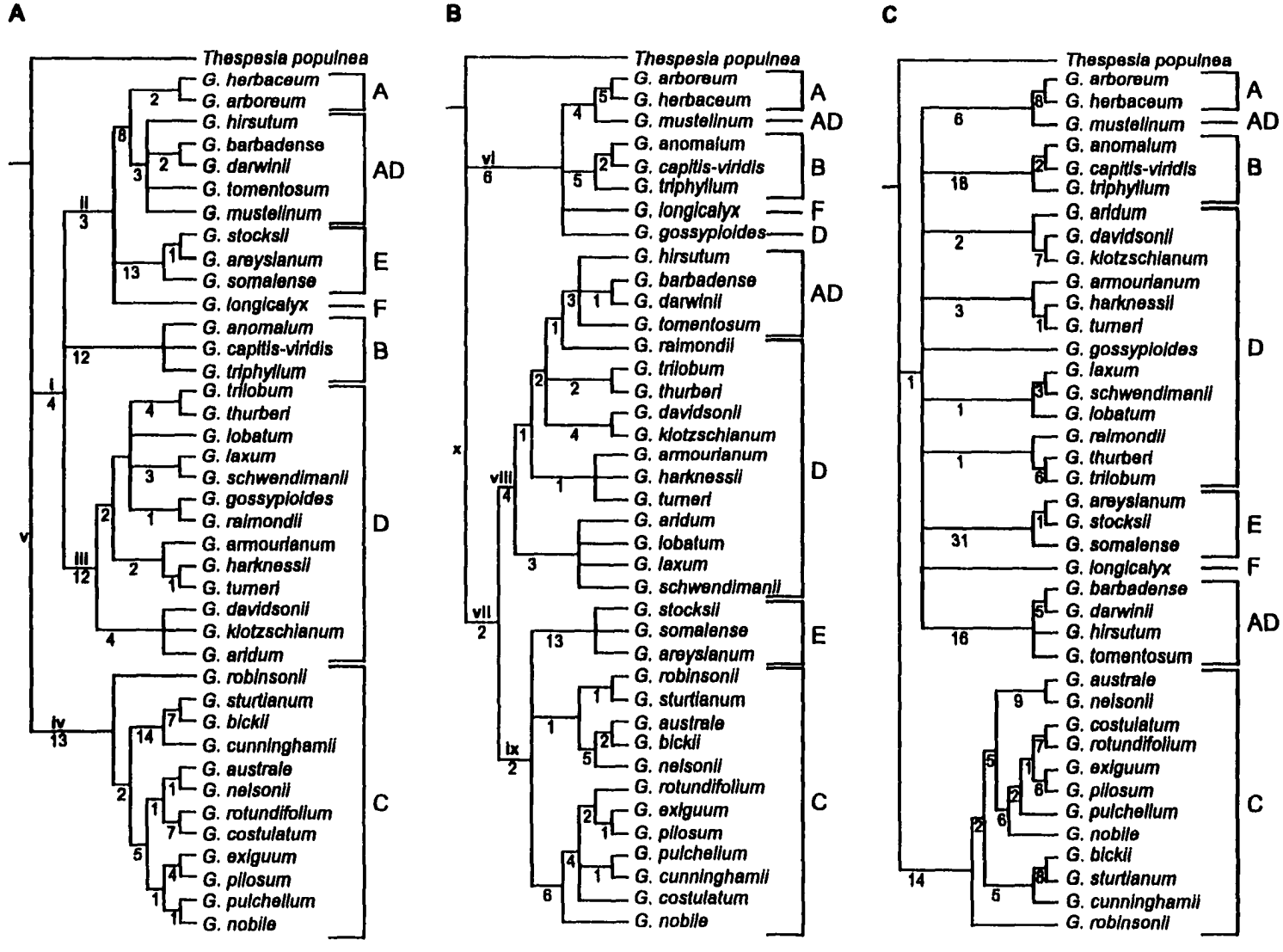

A
B
C
Thespesia populnea
G. herbaceum
G. arboreum
G. hirsutum
G. barbadense
G. darwinii
G. tomentosum
G. mustelinum
G. stocksii
G. areysianum
G. somalense
G. longicalyx
G. anomalum
G. capitis-viridis
G. triphyllum
G. trilobum
G. thurberi
G. lobatum
G. laxum
G. schwendimanii
G. gossypioides
G. raimondii
G. armourianum
G. harknessii
G. turneri
G. davidsonii
G. klotzschianum
G. aridum
G. robinsonii
G. sturtianum
G. bickii
G. cunninghamii
G. australe
G. nelsonii
G. rotundifolium
G. costulatum
G. exiguum
G. pilosum
G. pulchellum
G. nobile
A
AD
E
F
B
D
C

Fig. 7

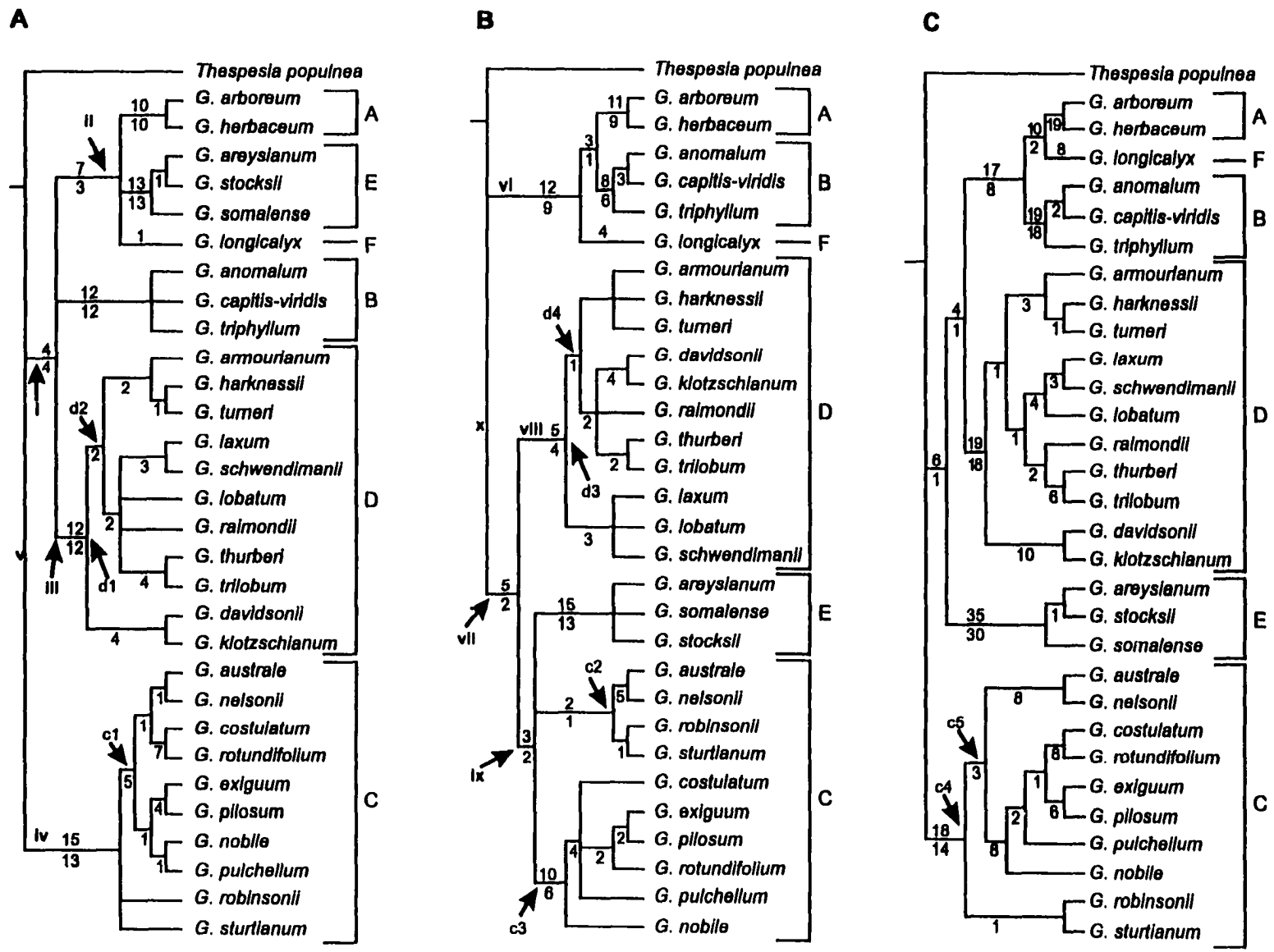

A
B
C
Thespesia populnea
G. arboreum
G. herbaceum
G. areysianum
G. stocksii
G. somalense
G. longicalyx
G. anomalum
G. capitis-viridis
G. triphyllum
G. armourianum
G. harknessii
G. tumeri
G. laxum
G. schwendimanii
G. lobatum
G. raimondii
G. thurberi
G. trilobum
G. davidsonii
G. klotzschianum
G. australe
G. nelsonii
G. costulatum
G. rotundifolium
G. exiguum
G. pilosum
G. nobile
G. pulchellum
G. robinsonii
G. sturtianum
A
B
C
D
E
F
I
II
III
IV
V
VI
VII
VIII
IX
X
d1
d2
d3
d4
c1
c2
c3
c4
c5

Fig. 8

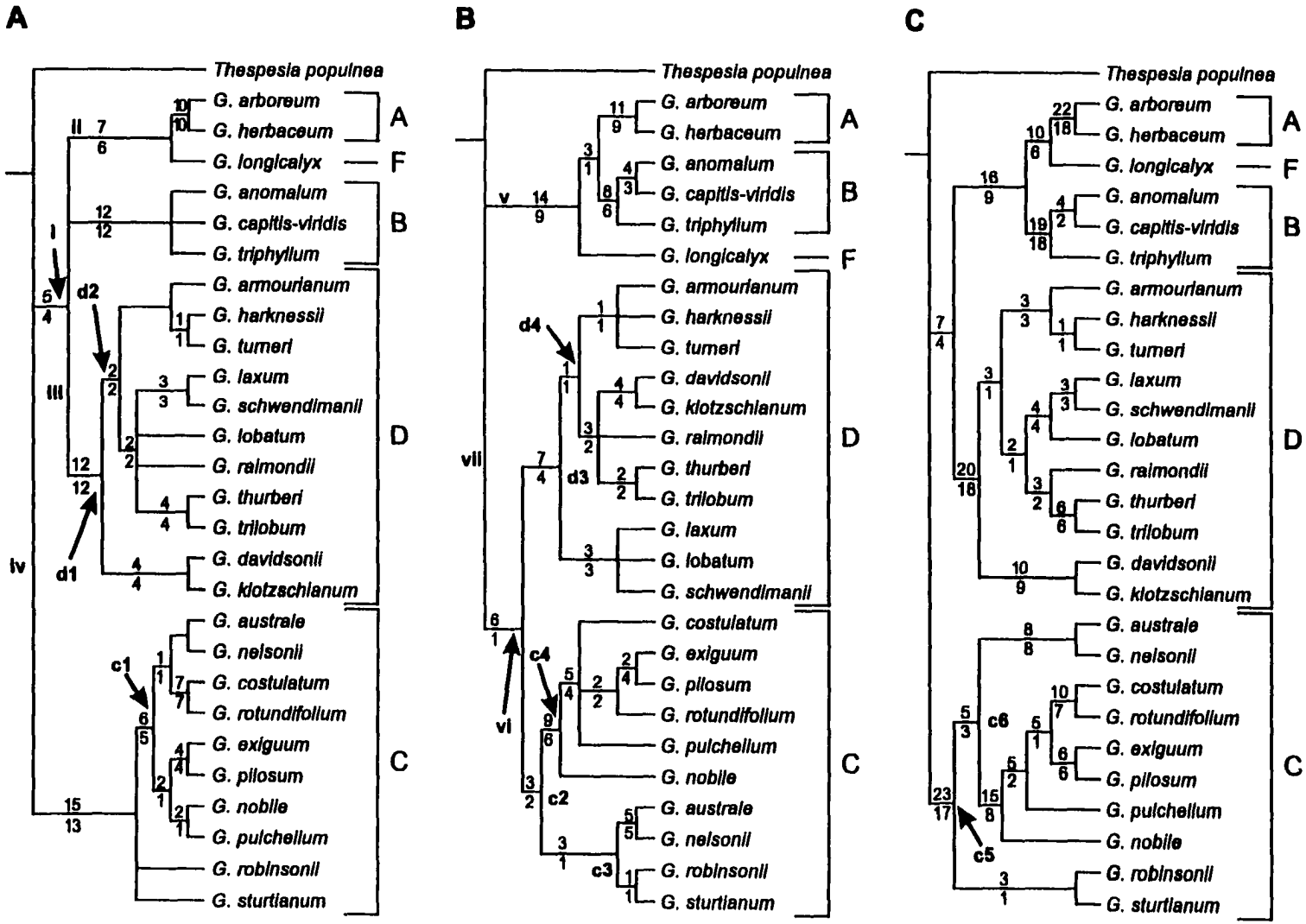

A
B
C
Thespesia populnea
G. arboreum
G. herbaceum
G. longicalyx
G. anomalum
G. capitis-viridis
G. triphyllum
G. armourianum
G. harknessii
G. turneri
G. laxum
G. schwendimanii
G. lobatum
G. raimondii
G. thurberi
G. trilobum
G. davidsonii
G. klotzschianum
G. australe
G. nelsonii
G. costulatum
G. rotundifolium
G. exiguum
G. pilosum
G. nobile
G. pulchellum
G. robinsonii
G. sturtianum
A
B
F
D
C
I
II
III
IV
V
VI
VII
d1
d2
d3
d4
c1
c2
c3
c4
c5
c6

Fig. 9

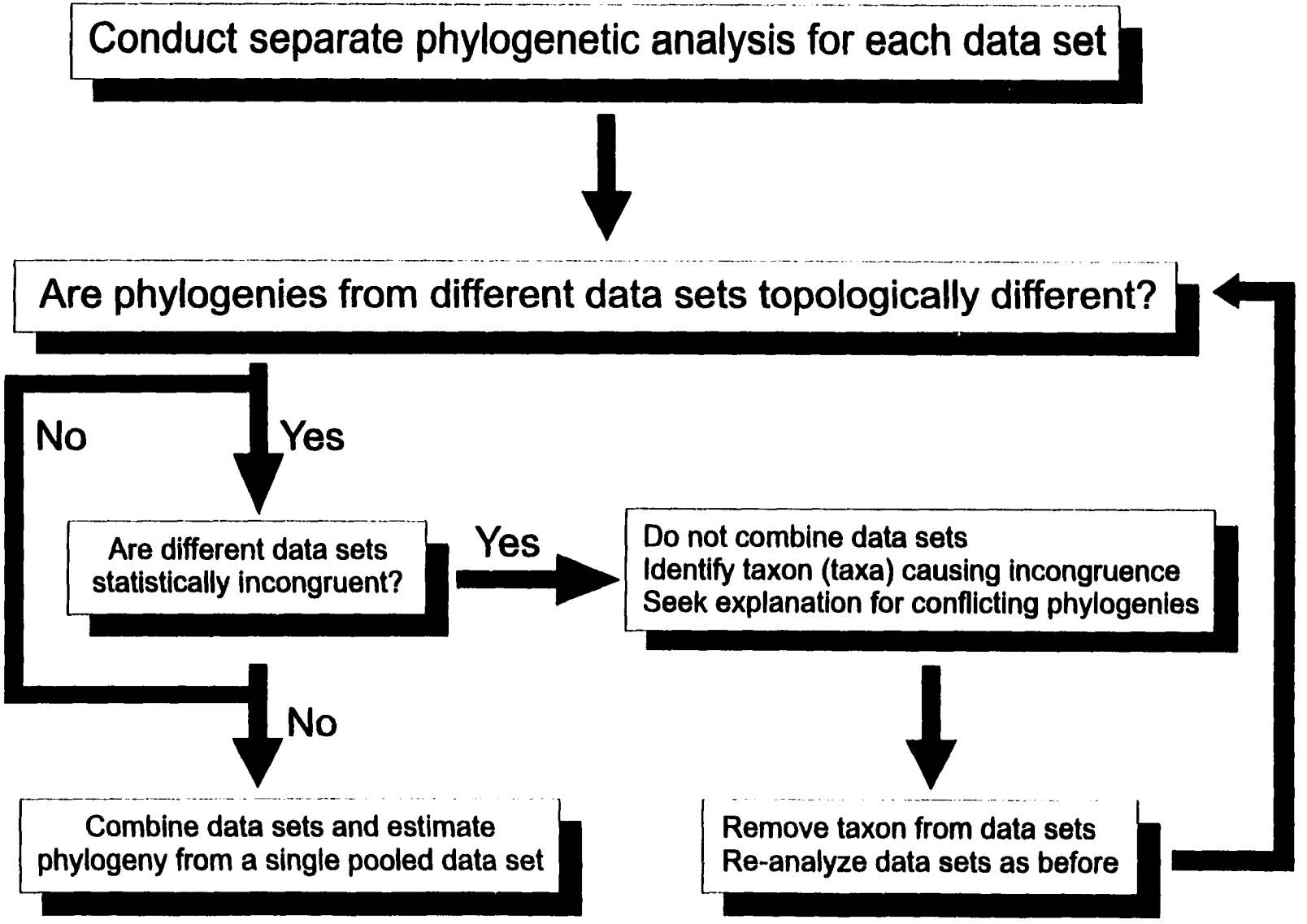

Conduct separate phylogenetic analysis for each data set
Are phylogenies from different data sets topologically different?
No
Yes
Are different data sets statistically incongruent?
Yes
Do not combine data sets
Identify taxon (taxa) causing incongruence
Seek explanation for conflicting phylogenies
No
Combine data sets and estimate phylogeny from a single pooled data set
Remove taxon from data sets
Re-analyze data sets as before

Fig. 10

CHAPTER 3. MOLECULAR SYSTEMATICS OF AUSTRALIAN *GOSSYPIUM* L. SECTION *GRANDICALYX* (MALVACEAE)

A paper to be submitted to the journal *Systematic Botany*

Tosak Seelanan[1,2], Curt L. Brubaker[3], James McD. Stewart[4], Lyn A. Craven[3], and Jonathan F. Wendel[5]

ABSTRACT

Australian *Gossypium* consists of 17 species classified into 3 sections. The largest and most poorly understood is the group of 12 species in section *Grandicalyx* that occur in the Kimberley and Cobourg regions of NW Australia. These areas are characterized by annual monsoon rains and dry-season fires. Species in sect. *Grandicalyx* exhibit a suite of morphological and ecological features that are otherwise unknown in the genus and that apparently evolved in response to seasonal fires and ant mutualism. These features include an herbaceous perennial habit, the ability to regenerate from thickened rootstocks following fires and extended drought, pedicels that recurve following pollination so that the capsules are pendent and open inverted at maturity, and sparsely vestitured, ant-dispersed seeds that bear elaiosomes. To better understand the evolutionary and biogeographic history of the species in sect. *Grandicalyx*, we sampled widely within and among species and generated three sets of DNA sequences, i.e., for the plastidial *rpl16* intron, and the nuclear 18S-26S ITS of rDNA and the alcohol dehydrogenase genes. Phylogenetic reconstructions with the three

---

[1] Graduate student, Department of Botany, Iowa State University

[2] Primary researcher and author

[3] Centre for Plant Biodiversity Research, CSIRO Division of Plant Industry, GPO Box 1600, Canberra, ACT 2601, Australia

[4] Department of Agronomy, 115 Plant Science Building, University of Arkansas, Fayetteville, AR 72701, U.S.A.

[5] Professor, Department of Botany, Bessey Hall, Iowa State University

data sets indicate that species in sect. *Grandicalyx* are monophyletic with little resolution within the clade. Sequence divergence is low in all pairwise comparisons among species, suggesting that the group radiated relatively recently, perhaps in the late Pliocene-Pleistocene after a much earlier (Miocene) divergence from the other extant Australian lineages. The palaeoclimatic record, sequence divergence estimates and phylogenetic data suggest that the diversification of the sect. *Grandicalyx* species arose via range fragmentation of a more widely distributed ancestor or ancestors. This evolutionary history was accompanied by the development of a prostrate to reclining herbaceous habit, adaptation to seasonal fires, and a suite of features associated with myrmecochory.

INTRODUCTION

Gossypium L. is a pan-tropical genus containing approximately 50 species of small trees, shrubs, and perennial herbs (Fryxell 1979, 1992; Stewart et al. 1997). The genus has been classified into four subgenera, of which one consists entirely of tetraploid taxa and the rest contain diploid species. Seven cytogenetically-defined genomic groupings (A through G and K) have been recognized (Endrizzi et al. 1985; Stewart 1995), of which four (A, B, E, and F; subgenus *Gossypium*) are African-Asian, one is primarily Mexican (D; subgenus *Houzingenia* Fryxell), and three (C, G, and K) are endemic to Australia. These latter species, comprising subgenus *Sturtia* (R. Br.) Tod., are the subject of the present investigation.

The Australian species have been classified into three sections (Fryxell 1979, 1992) that are congruent with the cytogenetic division of the subgenus into three genome groups. Section *Sturtia* (C genome) contains two species of sub-arborescent shrubs (1 - 4 m), *G. robinsonii* F. Mueller and *G. sturtianum* J. H. Willis. The former is confined to the Pilbara region of Western Australia, whereas the latter occurs in small isolated populations widely scattered across the sub-tropical to warm temperate arid zone of Australia, with populations in Queensland, New South Wales, South Australia, and Western Australia. Section *Hibiscoidea* Tod. (G genome) contains three species of shrubs (1-3 m) from the dry-monsoon and northern arid zone, with an aggregate geographic range that nearly spans the continent. The third section, sect. *Grandicalyx* (Fryxell) Fryxell, consists

entirely of herbaceous perennials, and is the most poorly understood. Seven of the 12 species presently recognized have been only recently described (Fryxell et al 1992; Stewart et al. 1997). Field observations suggest that the current taxonomy is inadequate, reflecting our incomplete understanding of this unique group of plants. All of the species are endemic to the Kimberley region of tropical NW Australia (Fig. 1) with the exception of one species restricted to the Cobourg peninsula on the central coast of the Northern Terrirtory; the former region lends its name to the genome of the group, which are designated by the prefix "K" (Stewart 1995). The geographic restriction of taxa in sect. *Grandicalyx* to the Kimberley region (except for *G. cunninghamii* Tod.) contrasts with the wider geographic range of species in the other sections (Fryxell, 1965; Craven et al. 1995).

Among Australian *Gossypium*, species in sections *Sturtia* and *Hibiscoidea* are more similar morphologically to the non-Australian *Gossypium* species than they are to the sect. *Grandicalyx* species. The former are upright shrubs to small trees with extrafloral nectaries on epicalices and leaves, and have seeds with a dense layer of pubescence. *Gossypium bickii* Prokh. possesses a few novel features such as a decumbent to suberect stature and a rootstock from which multiple shoots arise. In contrast, sect. *Grandicalyx* species are prostrate to upright herbaceous, multistemmed perennials, adapted to regenerate from the crown of a woody rootstock after the fires that frequent their seasonal monsoonal habitat. Species in sect. *Grandicalyx* typically have pedicels that recurve within 2-3 days of anthesis, so that capsule orientation is toward the ground, and in some species flowering and fruiting occur at ground level. In addition, there is a near-absence of the seed pubescence typical of most other *Gossypium* sections, and instead, seeds have elaiosomes that are half or more the length of the seeds (Fryxell et al. 1992). These morphological novelties have clearly evolved as adaptations for ant dispersal; upon capsule dehiscence, seeds fall passively to the ground where they are rapidly collected by ants. All authors involved in field collections (CLB, LAC, JMS, JFW) have observed ants to participate actively in the removal of seeds from the capsules of several of the species (also see Fryxell et al. 1992).

In addition to the morphological, geographical and ecological isolation of the group, sect. *Grandicalyx* species are cytogenetically distinct. Although all diploid species of *Gossypium* have the same chromosome number (n = 13), genome size varies considerably (Endrizzi et al. 1985), with species from Australia having the largest chromosomes. Within the Australian species, species in sect. *Sturtia* and sect. *Hibiscoidea* have 2C DNA contents of approximately 5 and 4.5 pg, respectively (Bennett et al. 1982), whereas the recently described species in sect. *Grandicalyx* (Fryxell et al. 1992; Stewart 1995; Stewart et al. 1997) have an estimated genome size of 7 pg (Stewart unpub.), the largest in the genus.

Previous phylogenetic studies in *Gossypium* (e.g., Wendel 1989; Wendel and Albert 1992; Wendel et al. 1995; Seelanan et al. 1997) focused on generic level relationships or groups other than Australian *Gossypium* (but see Wendel et al. 1991). Collecting expeditions to the Kimberley region in 1981, 1983, 1985 and 1993 have led to an increased understanding of the group and its morphological diversity (Fryxell et al. 1992; Craven et al. 1995; Stewart et al. 1997), while highlighting a number of taxonomic and phylogenetic questions. These trips also provided the quantity and diversity of accessions needed to address the evolution of the group. Morphological and molecular analyses of this material have allowed us to address the following questions: (1) Are the Kimberley cottons monophyletic, as presumed on the basis of their apparent morphological, ecological, and cytogenetic distinctiveness? (2) What are the phylogenetic relationships among K-genome species? (3) Can phylogenetic data be used in conjunction with other information to gain insight into the relative recency of origin of the presumptive clade, its phytogeographic history, and the evolution of its novel morphological features? These questions are addressed in the present study using inferences gained from DNA sequence analysis of three molecular data sets, one from the chloroplast genome (r*pl16* intron), and two from the nuclear genome (the internal transcribed spacer of the 18S-26S rDNA and an alcohol dehydrogenase gene).

MATERIALS AND METHODS

Plant Materials. Sixty-four accessions of 17 species of Australian *Gossypium* were used (Table 1). Leaf material for DNA extractions was

collected from individual plants growing in their native habitats or from first-generation plants grown from seeds obtained during collecting expeditions undertaken in 1981, 1983, 1985 and 1993. Voucher specimens are deposited at the Australian National Herbarium, Canberra, Australia, and at the Northern Territory Herbarium in Darwin, with many duplicates deposited at the Ada Hayden Herbarium at Iowa State University. Sampling density varied among species, and was primarily based on preliminary assessments of morphological diversity and relative breadth of geographic distribution. For species in sect. *Sturtia*, 5 accessions of the widely distributed *G. sturtianum* and 2 accessions of the more narrowly distributed *G. robinsonii* were sampled. For sect. *Hibiscoidea*, we sampled 2 accessions each of *G. nelsonii* Fryxell and *G. australe* F. Mueller and 3 accessions from *G. bickii* (one of an accession where leaves range from unlobed to slightly lobed, and two with leaves that are trilobed). Sampling of the K-genome taxa was as follows: 2 accessions of the restricted and geographically isolated *G. cunninghamii*; 3 accessions of each of the well-defined species *G. anapoides* J. M. Stewart, Craven & J. F. Wendel, *G. costulatum* Tod., *G. enthyle* Fryxell, Craven & J. M. Stewart, *G. londonderriense* Fryxell, Craven & J. M. Stewart, and *G. nobile* Fryxell, Craven & J. M. Stewart; and 4 to 6 (mostly) accessions from each of the more morphologically variable and in some cases intergrading species *G. exiguum* Fryxell, Craven & J. M. Stewart, *G. marchantii* Fryxell, Craven & J. M. Stewart (only 4), *G. pilosum* Fryxell, *G. populifolium*(Benth.) F. Mueller ex Tod., *G. pulchellum* (C. A. Gardner) Fryxell, and *G. rotundifolium* Fryxell, Craven & J. M. Stewart. Sampling localities are shown in Fig. 1.

Outgroups for Australian cottons as a whole (subgenus *Sturtia*) were selected based on cpDNA- and ITS-based phylogenies (Wendel and Albert 1992; Seelanan et al. 1997). These analyses indicate that the Australian group is monophyletic and sister to species having other genomes (A, B, D, E, F). For the present study, we selected the A-genome species, *G. arboreum* L. or *G. herbaceum* L., as outgroups largely because we have the complete sequences of *rpl16* intron and ITS from *G. arboreum* (Wendel et al. 1995; Small et al. 1998), and the sequence of *AdhD* from *G. herbaceum* (Small and Wendel, unpubl.). Our choice of outgroup taxa for different data sets should

have little effect on the resulting phylogenies because *G. arboreum* and *G. herbaceum* are closely related.

Data Acquisition. Total genomic DNA was extracted (Paterson et al. 1993) from silica-gel-dried leaf materials from field collections or from fresh leaves of plants grown from the original collections. Three regions of gene/intron/spacer were selected in this study: the *rpl16* intron from the chloroplast genome; and two sequences from the nuclear genome, the internal transcribed spacer (ITS) from the 18S-26S rDNA repeat and *AdhD*, one of several alcohol dehydrogenase genes present in diploid *Gossypium* species (Millar and Dennis 1996; Small et al. 1998; Small and Wendel, unpublished). Fig. 2 (top) shows the approximate locations of primer sites used to generate *AdhD* sequences. DNA sequence data were obtained by direct sequencing of double-stranded PCR templates.

Primers for double-stranded PCR amplifications of the *rpl16* intron are from Jordan et al. (1996). Amplification primers and PCR conditions for ITS are listed in Wendel et al. (1995). *AdhD* was amplified using the following primers: **X2-2** (5'-GCAATGGAGGTTCGTCTG-3') and **X8-3** (5'-GATCATGGCATTAATGTTTC-3'). These primers amplify a region from near the 3' end of exon 2 to the 5' end of exon 8. PCR amplifications for *AdhD* were performed in 50 μL reactions containing 1× *Taq* buffer, 2 mM of $MgCl_2$, 200 mM of each deoxynucleotide triphosphate, 0.2 mM of each primer, and 2.5 units of *Taq* DNA polymerase (Promega, Madison). Amplifications were carried out in a ThermoCycler PTC-100 (MJ Research) using the following parameters: 30 cycles of [1 min at 95C, 1 min at 50C, 2 min at 72C], followed by 7 min at 72C.

DNA sequencing was performed either manually or with automated DNA sequencing equipment. For the *rpl16* intron region, PCR products were cleaned using Wizard PCR Preps (Promega) and then sequenced using primer **F71** (Jordan et al. 1996), **F627** (5'-CGGAACAAACCAGAGACCAC-3'), **R699** (5'-TCGCGGGCGAATATTTAC-3'), and **R1516** (Kelchner and Wendel 1996). Sequences were obtained on an ABI DNA Sequencer using dRhodamine terminators; 300 to 400 ng of DNA template was used for each reaction. For ITS sequencing, PCR products (5 μL of 10+ ng/μL concentration) were treated with shrimp alkaline phosphatase and exonuclease (Amersham, Arlington Heights, IL), as suggested by the

manufacturer to remove unincorporated nucleotides and primers. Five μL of treated PCR products served as templates for DNA sequencing using the Thermosequenase terminator cycle sequencing system (Amersham) following the manufacturer's instructions. Due to GC compression of the ITS region in *Gossypium*, dITP was used in the termination mix. Electrophoresis was performed on 5% (long gel) and 6% (short gel) polyacrylamide gels (LongRanger, FMC) in 1× TTE (Tris-HCL/Taurine/ EDTA), followed by autoradiography for 2 - 3 days. For *AdhD*, PCR products were gel-fractionated in 1.5% low-melt agarose (Promega), stained with ethidium bromide, and visualized using UV light. PCR products having the correct size were removed from the gel and used as templates for DNA sequencing. DNA sequencing, electrophoresis, and autoradiography were the same as for ITS except that dGTP was used in the termination mix. In a few cases, dITP was used to resolve compression. Primers used in *AdhD* sequencing were the amplification primers **X2-2** and **X8-3**, and the internal primers **X6** (5'-ATCAACACCAATAATCCTAGAA-3'), and **X5-1** (5'-GCCACAGTTGAACCTTTG-3'). In some cases, it was necessary to employ the additional internal primers **X5-2** (5'-AATAATTTTCGAGGTCTTGG-3') and **X4-4** (5'-ACCTCACCCACACTCTCAAC-3').

Copy number estimation for AdhD. Southern blot analysis was employed to verify that *AdhD* is a single-copy gene. Representative accessions from each section were used, including *G. sturtianum* (sect. *Sturtia*), *G. bickii* (sect. *Hibiscoidea*), and *G. costulatum*, *G. exiguum*, *G. londonderriense*, and *G. pulchellum* (sect. *Grandicalyx*). Probe selection for Southern blot analysis was determined from comparison of sequences of all *Adh* loci. Preliminary experiments (Small and Wendel, unpubl.) indicated that the intron 3 region would be useful as an *AdhD* locus-specific probe. Choice of restriction endonucleases was determined from the complete *AdhD* sequences of ingroup taxa, and two unique restriction sites, one upstream and the other downstream of intron 3, were selected. Approximately 10 μg of each genomic DNA was digested with each of two restriction endonucleases, *Nsi*I and *Eco*RI. Agarose gel electrophoresis, transfer to nylon membranes, hybridization of membrane-bound fragments to [^{32}P]dCTP-labeled probes, and autoradiography were performed as

described in Sambrook et al. (1989). *Hin*dIII/*Eco*RI-digested l DNA fragments were used as size standards.

Data Analysis. For all three data sets, character-based phylogenetic analyses were performed using PAUP 3.1.1 (Swofford 1993) and MacClade 3.07 (Sinauer, Sunderland, Massachusetts), and distance based phylogenetic analyses were carried out using PHYLIP version 3.572 (PHYLogeny Inference Package, Department of Genetics, University of Washington, WA). MEGA 1.1 (Kumar et al. 1993) was used for statistical characterization of each data set. In character-based phylogenetic analyses, phylogenetic signal in each data set was determined as suggested by Hillis and Huelsenbeck (1992). We estimated g_1 values by sampling 10^6 random trees in PAUP. Tree estimation was carried out by using the following parameters in PAUP: HEURISTIC search with 1,000 random sequence additions, 10 trees kept in each replicate, MULPARS ON and TBR branch swapping. Bremer support (Bremer 1988) was evaluated as described elsewhere (Seelanan et al. 1997). Data sets were analyzed as follows:

rpl16 *intron.* Sequences were aligned initially using GCG—pileup and then manually adjusted as necessary. Alignment was relatively straightforward given the small number of problematic gaps. Regions where sequence alignment was questionable were excluded from the analyses. Because *rpl16* intron sequences from accessions within the same species showed little to no variation, redundant accessions were omitted from the analyses. Depending on how potentially phylogenetically informative gaps were coded, four different analyses were conducted; namely, (a) gaps were coded as missing; (b) gaps were coded as presence/absence, but gaps at the original position in the aligned sequences were excluded from the analysis; (c) gaps were excluded with no re-coding; and (d) gaps were coded as presence/absence characters and the original gaps were included in the analysis and coded as missing.

ITS. Sequence alignment required only a few one-bp gaps to obtain proper alignment and was done by hand. Phylogenetic analyses of ITS data were the same as those performed for *rpl16* except that all gaps were treated as missing data, and no gaps were scored as presence/absence characters. Ribosomal DNA (rDNA) genes are biparentally inherited and exist in highly reiterated tandem arrays. Consequently, when PCR products are used for

DNA sequencing, low levels of sequence polymorphism may occur, either from allelic heterozygosity or incomplete homogenization by concerted evolution. In such cases, polymorphisms (referred to MSTAXA in PAUP 3.1.1) were treated as UNCERTAINTY or POLYMORPH when analyzed in PAUP 3.1.1. Preliminary analysis showed that trees recovered when polymorphism was treated as UNCERTAINTY were the same as or a subset of those recovered when polymorphism was treated as POLYMORPH. However, measures of consistency, homoplasy, and tree length differed considerably in the two analyses. Since there was no difference in the tree topology regardless of polymorphism treatment, polymorphisms were treated as UNCERTAINTY. For analysis using MEGA, polymorphisms were treated as missing data, because MEGA cannot handle polymorphic characters.

AdhD. Sequences of *AdhD* were easily aligned, requiring only a few one- to several-bp gaps. Three phylogenetic analyses of *AdhD* were carried out: (1) for the whole data set; (2) using only data from exon sequences; and (3) using only intron sequences. Gaps were treated as missing data with the exception that one indel (insertion/deletion) in intron 5 was excluded. Polymorphism was anticipated if plants were heterozygous at *AdhD*, and indeed, some polymorphism was detected. In these cases the polymorphic characters were treated as UNCERTAINTY for all analyses (or missing data using MEGA).

RESULTS

Sequence Variation

rpl16 *intron*. The aligned matrix of *rpl16* intron sequences was 1,155 bp in length. This matrix starts at the twentieth nucleotide of the *rpl16* intron and ends at the seventeenth nucleotide of exon 2. Thus, nearly the entire length of each sequence (1,138 bp of aligned sequence) was composed of introns. Overall, sequences from ingroup taxa were easily alignable, requiring only insertions of gaps of only one to several bp. In one region 62 bp in length (corresponding to positions 455-517 of the alignment), four indels occurred, with one instance (in *G. anapoides*) representing an insertion of a direct repeat 7 bp in length (corresponding to positions 455-517). When aligning sequences from the outgroups, this

same region of 62 bp was unalignable to the corresponding region in the ingroup. Accordingly, this region was not included in any of the phylogenetic analyses. Although *rpl16* sequences from multiple accessions of the same species were often identical, intraspecific variation was found in sequences from *G. sturtianum*, *G. bickii*, and 10 of the 12 species of sect. *Grandicalyx*. On average, *rpl16* sequences have a GC content of 30%, as is typical of this region in other plants (Palmer 1991). Of the 1,138 bp used for phylogenetic analysis, 26 sites were variable within the data set as a whole. For ingroup taxa alone, 20 sites were variable, and only 9 sites were potentially phylogenetically informative, excluding gaps. There were 15 indels, of which 5 were potentially phylogenetically informative. Because 2 potentially informative indels were within the excluded region (see above), only 3 indels were included in phylogenetic analyses.

ITS. Of 64 ingroup taxa, only 61 ITS sequences were obtained. ITS sequences could not be obtained from two accessions of *G. nobile* [4229 and 4235] and one accession of *G. enthyle* [9224]. As a result, only one ITS sequence from *G. nobile* [4231] and two sequences from *G. enthyle* [9238 and 9239] were included. Aligned ITS sequences were 688 bp in length (ITS1 = 296 bp; ITS2 = 228 bp), with the alignment requiring only seven single-bp indels. The 5.8S gene was 164 bp in all taxa. Four and three single-bp indels were found in ITS1 and ITS2, respectively. The GC content of ITS, on average, was double that of the *rpl16* intron, or 60%. Of 688 aligned nucleotides, 88 were variable: 56 in ITS1, 1 in 5.8S, and 31 in ITS2. When considering only ingroup taxa, 60 sites were variable, of which 39 were potentially phylogenetically informative: 21 in ITS1, 1 in 5.8S, and 17 in ITS2.

AdhD. Based on Southern blot experiments (Fig. 2) of representative taxa from each section, we inferred that *AdhD* occurs as a single locus in Australian *Gossypium*. This increased our confidence that we were sampling orthologous sequences from the accessions and species studied. Sequences of *AdhD* started in exon 2 and ended in exon 8 (Fig. 2). Aligned sequences of *AdhD*, excluding an unreadable region in intron 3, were 1,601 nucleotides in length, consisting of 684 bp of exon sequence and 917 bp of intron sequence. No indels were detected in any of the seven exons sequenced, but 18 gaps were required in the introns. Each of these was

only one to several bp in length. Of the 18 indels, one (in intron 5) that occurred exclusively in all species of sect. *Grandicalyx* appeared to be caused by a series of mutational events, such as an inversion in conjunction with indels in flanking regions. Accordingly, this indel was excluded from the analyses. Of the indels detected, four were potentially phylogenetically informative, of which one appeared to be homoplasious. For the whole data set, 136 sites — 39 and 97 in exon and intron positions, respectively — were variable. When considering only ingroup taxa, 114 sites (33 exon, 81 intron) were variable. Of these 114 sites, 42 (13 exon, 29 intron) were potentially phylogenetically informative.

Table 2 summarizes some of the characteristics of the *AdhD* sequences sampled. *AdhD* sequences were partitioned into exon and intron regions, revealing differences in GC content (~50% vs. ~30% in exons vs. introns, respectively), and amounts of divergence as indicated by the number of nucleotide substitutions in exon and intron regions from species in different sections. As illustrated in Table 2, divergence levels for exons of species in sect. *Sturtia* and sect. *Hibiscoidea* are lower than those for introns. In contrast, there is no difference in divergence amount for exons and introns of species in sect. *Grandicalyx*. In addition, divergence levels for exons in sect. *Sturtia* are about half of those in sect. *Hibiscoidea*, whereas the amount of divergence is the same for introns. Finally, synonymous substitutions predominate in all three sections: synonymous substitution rates are 6.5, 3.0, and 18.5 times higher than nonsynonymous substitution rates in sect. *Sturtia*, sect. *Hibiscoidea*, and sect. *Grandicalyx*, respectively.

Phylogenetic Analysis

rpl16 *intron*. For this data set, the g_1 statistic was -1.20, which indicates that the data are more structured than random ($P > 0.01$). Four separate analyses were carried out. In the first analysis, gaps were treated as missing data. Only one most parsimonious tree with no homoplasy was recovered (Fig. 3, Tree **1**), with a length of 26 steps (9 steps if uninformative characters are excluded). This tree consists of an unresolved trichotomy, with the three branches being *G. robinsonii*; *G. sturtianum*, *G. bickii*, and *G.*

cunninghamii; and all remaining taxa. There was little resolution in the last clade.

In the second analysis, gaps were excluded from the data matrix, but recoded as presence/absence characters. Six most parsimonious trees were found, and none was the same tree as that in the first analysis. Trees were 35 steps in length with a CI of 0.971 and a RI of 0.981 (tree length = 13 and CI = 0.923 when uninformative characters are excluded). The topologies of the most parsimonious trees found in this second analysis were similar to Tree **1** in Fig. 3, with the replacement of a topology in the dashed box with one of topologies **A**, **B**, **C**, or **D** in the inset. It should be noted that although there were 6 trees found in this analysis there were only 4 sets of trees: the first set consisted of 2 trees whose replacement topology is **A**; the second set also had 2 trees whose replacement topology was the same as topology **B**; the third and the fourth sets each contained only one tree whose replacement topologies were **C** and **D**, respectively. Inspection of both trees in the first and the second sets revealed that the difference between the two trees in a given set was due to the zero-in-length branch assigned by PAUP for the (*G. bickii-G. cunninghamii-G. sturtianum*) clade in topology **A**, and for the (*Gossypium bickii-G. robinsonii-G. sturtianum*) clade in topology **B**. Specifically, in topology **A**, PAUP assigned the clade of these 3 species as either ((*G. cunninghamii* 499785, *G. cunninghamii* 499785), (*G. bickii* 464843, (*G. bickii* 9360, *G. sturtianum* 499810, *G. sturtianum* 5237))) or ((*G. cunninghamii* 499785, *G. cunninghamii* 499785), *G. bickii* 464843, (*G. bickii* 9360, *G. sturtianum* 499810, *G. sturtianum* 5237)). In topology **B**, PAUP assigned the clade of these 3 species as either (*G. robinsonii* 9297, (*G. bickii* 464843, (*G. bickii* 9360, *G. sturtianum* 499810, *G. sturtianum* 5237))) or (*G. robinsonii* 9297, *G. bickii* 464843, (*G. bickii* 9360, *G. sturtianum* 499810, *G. sturtianum* 5237)). When the zero-in-length branches were collapsed, only topologies **A** and **B** were found. As a result, only four replacement topologies are shown in Fig. 3. The differences among these six trees were that: (a) inset taxa formed either a dichotomy or trichotomy; (b) when the inset taxa formed a dichotomy, there were two sets of trees: one set consisted of trees where *G. robinsonii* was sister to the *G. sturtianum-G. bickii-G. cunninghamii* clade (Tree **1** with topology **A**), and the other one contained trees where *G. cunninghamii* was sister to the *G. sturtianum-G.*

bickii-G. robinsonii clade (Tree **1** with topology **B**); (c) when inset taxa formed a trichotomy, either *G. robinsonii* was in a separate clade and *G. cunninghamii* was in the *G. sturtianum-G. bickii* clade (Tree **1** with topology **C**), or *G. robinsonii* was in the *G. sturtianum-G. bickii* clade but *G. cunninghamii* was in a separate clade (Tree **1** with topology **D**). In all cases, the clade consisting of *G. australe*, *G. nelsonii*, and the rest of species in sect. *Grandicalyx* remained topologically unchanged. This unchanged clade had the same topology as the respective clade in the tree from the first analysis.

In the third analysis, gaps were excluded and not recoded. This procedure led to the recovery of only one most parsimonious tree, with tree statistics and a topology that were identical to those resulting from the first analysis. In the fourth and final analysis, gaps were treated as missing and were also recoded as presence/absence characters. Six most parsimonious trees were recovered, identical to the trees recovered in the second analysis.

A neighbor-joining analysis, using distances based on Jukes-Cantor corrections for multiple hits (Jukes and Cantor 1969), is shown in Figure 3 as Tree **2**. This tree shows the same primary three clades as recovered in the parsimony analysis, a similar absence of resolution within the clade consisting of all taxa but *G. cunninghamii* of sect. *Grandicalyx* and *G. australe* and *G. nelsonii*, and a relationship among accessions shown in the inset that was not found in any of the character-based analyses.

ITS. Due to the limitation on the number of maximum trees that may be retained in PAUP, we adopted the following procedure to estimate the most parsimonious trees from the ITS data as well as from the *AdhD* data set: searches were conducted as described in ***Materials and Methods*** but at least ten runs were carried out. Trees from each run were compared, and it was found that varying percentages of trees were the same in separate runs. The absence of complete intersection of the sets of trees recovered by independent runs indicated that each run explored slightly different islands of trees. Because the differences among the tree islands appeared to be minor, we compared the strict consensus tree from each run. In every case, each run gave the same strict consensus tree. While this procedure does not guarantee that the single shortest tree or set of trees has been found, this procedure, combined with the limited sequence variation involved in the

analyses, suggested that the resulting estimated phylogenies represented reasonable approximations of what could possibly be inferred from parsimony analysis of the data at hand. Because so many equally shortest trees were found for both the ITS and *AdhD* data sets, we present only the strict consensus tree for each.

For the ITS data set, the g_1 statistic was -1.17, which indicates that the data are more structured than random ($P > 0.01$). The strict consensus is shown in Fig. 4 (Tree **1**). Tree lengths for the constituent trees are 106, with CI and RI values of 0.887 and 0.947, respectively (tree length = 67 and CI = 0.821 if uninformative characters are excluded). The strict consensus tree indicates that the clade consisting of species in sect. *Sturtia* and sect. *Hisbiscoidea* (*G. robinsonii* through *G. australe*, top) is sister to the clade of species in sect. *Grandicalyx* (all others, except the outgroup species *G. arboreum*). Within the former clade, species in sect. *Sturtia* (*G. robinsonii* and *G. sturtianum*) are sister to sect. *Hibiscoidea* (*G. australe*, *G. bickii*, and *G. nelsonii*), although resolution regarding the relationship among the three species in sect. *Hibiscoidea* is unclear. In the clade consisting of all species in sect *Grandicalyx*, there was strikingly little resolution. A few weakly supported clades were formed: the clade of *G. exiguum*, *G. londonderriense*, *G. pilosum*, *G. rotundifolium* and one accession of *G. marchantii*; the clade of *G. anapoides*, *G. cunninghamii*, and another accession of *G. marchantii*; the clade of *G. populifolium* (accession 9203) and *G. costulatum* (accession 9260); and the clade containing all sect. *Grandicalyx* taxa but *G. nobile*.

Distance-based phylogenetic analysis led to the recovery of Tree **2** in Figure 4. This tree is similar to that obtained from parsimony analysis in most respects, including the resolution of a monophyletic sect. *Grandicalyx* and resolution of a (sect. *Sturtia* + sect. *Hisbiscoidea*) clade as the sister group of sect. *Grandicalyx*. The neighbor-joining tree highlights the relative paucity of ITS sequence evolution within the clade of sect. *Grandicalyx*, as indicated by the minuscule branch lengths shown. Finally, the neighbor joining and parsimony analyses yield different inferences regarding the position of *G. robinsonii* and the relationships among the sect. *Hibicoidea* species.

AdhD. Values of the g_1 statistic for the entire *AdhD* sequence (exon and intron), exons alone, and introns alone were -1.52, -1.38, and -1.33,

respectively. This indicates that data from each partition is more structured than random ($P > 0.01$). Separate analyses were performed on each of these three partitions, and the results are summarized as shown in Fig. 5 (Tree **1**), which presents a strict consensus obtained for the entire sequence, exon partition, and intron partition, respectively. When analyzing the entire *AdhD* region, the shortest trees had 146 steps with CI and RI values of 0.959 and 0.964, respectively (tree length = 69 and CI = 0.913 when uninformative characters were excluded). Inferences based on the phylogeny estimated from the entire *AdhD* data set include (1) species of sect. *Grandicalyx* are monophyletic, with high branch support; (2) within this clade, all accessions of *G. exiguum*, *G. pilosum*, *G. rotundifolium* and one accession from each of *G. enthyle*, *G. marchantii*, and *G. populifolium* form a weakly supported monophyletic group; (3) all accessions of *G. londonderriense* form a monophyletic group with one accession of *G. costulatum*; (4) *G. cunninghamii* is sister to *G. anapoides*; (5) sect. *Sturtia* is monophyletic; and (6) sect. *Hibiscoidea* is not monophyletic.

When only exon nucleotide positions were included in the analysis, the shortest trees had 40 steps with no homoplasy (tree length = 18 and CI=1.0 when uninformative characters were excluded). The strict consensus tree and inferences from this analysis were nearly the same as those found when the entire *AdhD* region was analyzed, as shown in Fig. 5. The differences between the two consensus trees were that: (1) branches supporting *G. londonderriense* and *G. costulatum* collapsed; and (2) the branch uniting *G. cunninghamii* and *G. anapoides* disappeared.

When phylogenetic trees were estimated from only intron positions, the shortest trees obtained had 106 steps, with CI and RI values of 0.943 and 0.945, respectively (tree length = 51 and CI = 0.882 when uninformative characters were excluded). The strict consensus tree from this analysis is similar to that from analysis of the entire data set, with the following exceptions: (1) sect. *Sturtia* species were not resolved as monophyletic (thick line, Fig. 5, Tree **1**); and (2) the branch holding all accessions of *G. exiguum*, *G. pilosum*, *G. rotundifolium* and one accession from each of *G. enthyle*, *G. marchantii*, and *G. populifolium* was not present.

Neighbor-joining analyses of *AdhD* were conducted in the same manner as the parsimony analyses, namely, with separate runs for three

partitions of the data. Results from each analysis paralleled those obtained from the character-based analyses. The tree obtained from the entire *AdhD* region is shown in Fig. 5 (Tree **2**). Overall, the tree topology leads to the same inferences as those obtained from the parsimony analysis. As was the case for the *rpl16* and ITS data, branch lengths within the clade of sect. *Grandicalyx* are short, once more underscoring the limited sequence evolution that has occurred among sect. *Grandicalyx* species.

DISCUSSION

Phylogenetic inferences

Inferences based on cpDNA sequences. Previous phylogenetic analyses that included Australian cottons (Wendel et al. 1991; Wendel and Albert 1992; Seelanan et al. 1997) sampled few accessions from the Kimberley region. Nonetheless, these earlier studies suggested several cladistic relationships that were also found in the present study. Of particular note is the phylogeny inferred from the chloroplast *rpl16* intron (Fig. 3), which corroborates the close cytoplasmic affinity of *G. bickii* and *G. cunninghamii*, species from sect. *Hibiscoidea* and sect. *Grandicalyx*, respectively, to each other and to *G. sturtianum*, a species in sect. *Sturtia*. These results conflict with data from morphology and from nuclear molecular markers, and hence, inclusion of these three species, one from each of the three taxonomic sections, in a single clade has been interpreted as reflecting past hybridization between their ancestors (Wendel et al. 1991; Wendel and Albert 1992). Because of this past reticulation, sect. *Sturtia*, sect. *Hibiscoidea*, and sect. *Grandicalyx* are paraphyletic, when considering data from the chloroplast genome.

A second inference from the *rpl16* data, also reported by Wendel and Albert (1992), is the inclusion of the other two species of sect. *Hibiscoidea*, *G. australe* and *G. nelsonii*, with the remaining species of sect. *Grandicalyx*. However, no inferences regarding the relationship of these two species in sect. *Hibiscoidea* to any species in sect. *Grandicalyx* can be drawn from the character or distance based phylogenies based on the *rpl16* data (Fig. 3). As is evident from the lack of character support in the parsimony-based tree and the short branches in the NJ tree (Fig. 3), *rpl16* possessed little variation (26 variable nucleotide positions, of 1,155 scored) in this group of

taxa, and hence few potentially informative characters (9) were available for phylogeny reconstruction.

We view the near-absence of *rpl16* sequence evolution within the diverse assemblage of species in sect. *Grandicalyx* (and those two species in sect. *Hibiscoidea*) to be a notable result in itself. The *rpl16* intron is often selected for use in phylogenetic studies because of its relatively high rate of sequence evolution (Kelchner and Wendel 1996; Kelchner and Clark 1997; Small et al. 1998; Schnabel and Wendel 1998), yet in the present study only two nucleotide substitutions were detected that serve as synapomorphies within sect. *Grandicalyx* (Tree **1**, Fig. 3). The remaining seven potentially informative characters support branches uniting *G. sturtianum-G. bickii-G. cunninghamii* (1 character), branches within that clade (4 characters), and the branch consisting of sect. *Grandicalyx* taxa along with *G. australe* and *G. nelsonii* (2 characters).

In other plants the *rpl16* intron has been found to experience relatively rapid rates of insertions and deletions, and in some cases these indels have proved to be useful phylogenetic characters (for example, in the bamboo genus *Chusquea*; Kelchner and Clark 1997). In the present study, 15 indels were detected, and of these 11 were included in phylogenetic analyses (four were excluded due to alignment uncertainties). Nine of the 11 indels were autapomorphic, while the remaining two provide the only characters that unite *G. australe*, *G. nelsonii*, and all but *G. cunninghamii* together.

Inferences based on sequences from the nuclear genome. The limited sequence evolution observed with *rpl16* was paralleled in the ITS and *AdhD* data sets, yet they concur with respect to several phylogenetic inferences. First, sect. *Grandicalyx* is a monophyletic group, as indicated by the high support for monophyly in both data sets (11 characters with Bremer support of 10 for ITS; 14 characters with Bremer support of 13 for *AdhD*). Second, in contrast to the results based on cpDNA sequences, nuclear phylogenies lead to the inference that sect. *Sturtia* is monophyletic. This inference is strongly supported by the ITS data (6 characters with Bremer support of 6), but is only weakly (NJ analysis) or not at all (parsimony analysis) supported by the *AdhD* data. A third result is the revelation of a close relationship between the newly (and as yet informally)

described (Stewart et al. 1997) species *G. anapoides* and the geographically disjunct and cytoplasmically introgressant (Wendel and Albert 1992) species *G. cunninghamii*. This close relationship has low (1 character) but consistent character support across data sets. In the *AdhD* data, support consisted of a synapomorphic deletion of two nucleotides (scored as a single character) near the 3' end of intron 7, and hence, this clade does not appear in phylogenies generated from exon data alone (Fig. 5). Fourth, and finally, both the ITS and the *AdhD* data sets revealed virtually no resolution within sect. *Grandicalyx*, reflecting the paucity of molecular evolution that has occurred in these taxa since their divergence from a common ancestor. Although both data sets provide as many as 42-44 potentially phylogenetically informative characters overall, one-half to two-thirds of these characters support monophyly of the three sections, leaving relatively few characters for inferring relationships among taxa within each section. With respect to sect. *Grandicalyx*, there were far more terminals included in the analyses (32 for ITS; 42 for *AdhD*) than there were potentially phylogenetically informative characters (10-20), and hence, lack of resolution was unavoidable.

Given the low level of resolution within the sect. *Grandicalyx* clade, there were few discrepancies between the ITS and *AdhD*-based phylogenies. One discrepancy involved relationships among sections, where in the ITS-based phylogeny, sect. *Sturtia* is sister to sect. *Hibiscoidea*, and the clade of these clades is sister to sect. *Grandicalyx*. For the *AdhD* data, this hierarchy is only recovered using NJ analysis, although results from parsimony analysis do not directly contradict this inference (Fig. 5). Support for the sister-group relationship between sect. *Sturtia* and sect. *Hibiscoidea* in the ITS data set is low (3 characters with Bremer support of 1). A second difference between the *AdhD* and ITS phylogenies concerns relationships among the three species in sect. *Hibiscoidea*: *G. australe*, *G. bickii*, and *G. nelsonii*. These three form a strongly supported monophyletic group based on ITS data, whereas for *AdhD* this inference is supported only by the NJ analysis. A third discrepancy between the topologies based on the *AdhD* and ITS data sets concerns the inclusivity of the "prostrate" clade, i.e., the weakly but consistently supported clade that includes the prostrate to decumbent taxa *G. exiguum*, *G. rotundifolium*, and *G. pilosum*. The two

phylogenies differ with respect to the inclusion of *G. londonderriense*, *G. populifolium* accessions 4257 and 499808, and *G. enthyle* accession LC9239.

Comparison of phylogenies based on cpDNA and nuclear sequences. Comparisons of the phylogenetic results of the three molecular data sets permit several conclusions. First, sect. *Grandicalyx* appears monophyletic, in accordance with expectations based on their many morphological, ecological, cytogenetic and geographical distinctions. The single exception to molecular monophyly concerns the aberrant species *G. cunninghamii*, which appears to have experienced cytoplasmic introgression from a species in sect. *Sturtia* (Wendel and Albert 1992; present study, Fig. 3 vs. Figs. 4, 5). Identification of *G. anapoides* as the closest relative of *G. cunninghamii*, in the *AdhD* and ITS data sets, suggests a possible lineage of the paternal ancestor of the putatively reticulate species; this suggestion is a second noteworthy inference based on the phylogenetic results. A third conclusion is that species in sect. *Grandicalyx* appear to have diversified relatively recently (as discussed below). In this respect, perhaps the most impressive feature of the three molecular data sets is that species in sect. *Grandicalyx* exhibit so little sequence variation. Nearly 2 kb of nuclear sequence (two-thirds intron and spacer) and over 1 kb of non-coding cpDNA sequence was generated for each accession, and yet virtually no resolution within sect. *Grandicalyx* was obtained using any of the three data sets. Because of the absence of resolution there was no opportunity for incongruence between data sets, and for this reason, as well as because sampling of accessions was not strictly parallel across data sets, we did not perform any combined analyses (*ala* Seelanan et al. 1997). Fourth, the molecular data suggest that both sect. *Sturtia* and sect. *Hibiscoidea* are monophyletic, although the two species of the former are paraphyletic in the cpDNA tree (Fig. 3) and the three species in the latter remained unresolved in the tree resulting from parsimony analysis of the *AdhD* data (Fig. 5). Finally, the molecular data identify, if only incompletely, a clade consisting of the prostrate Kimberley species of sect. *Grandicalyx*.

Intraspecific polymorphism

The low levels of sequence variation observed in all three molecular data sets suggest that there should be little to no variation among accessions of most species of sect. *Grandicalyx*. This expectation is bolstered by the fact that many of the species occur in small populations in highly restricted geographic ranges. Nonetheless, consistent but low levels of variation were found within most species. Multiple accessions were sampled from most species and as few as two accessions were sampled from others. For the *rpl16* intron data, 12 of 17 species sampled showed intraspecific variation (nucleotide substitutions and indels). Of these 12 species, 10 are in sect. *Grandicalyx*, with one each from sect. *Sturtia* and sect. *Hibiscoidea*. For the ITS data, all species sampled exhibited intraspecific variation, whereas for *AdhD*, all species except *G. cunninghamii* showed intraspecific variation. Most polymorphisms within species are autapomorphic, and hence do not impact our phylogenetic inferences. Several polymorphisms are shared among accessions of the same species or represent shared polymorphisms among species. The latter may have arisen from a number of causes, including sequencing error, lineage sorting, and interspecific gene flow. Perhaps the most interesting polymorphisms are those that occur across species boundaries, i.e., the same polymorphism at a particular site is 'shared' by more than one species. For example, one polymorphic *AdhD* site is shared by *G. pilosum* [accessions 499785, 499786, 499803, and CampRd], *G. populifolium* [accession 499808], and *G. rotundifolium* [accessions 5217, 9281, and 499789], which is notable in that these accessions are in close geographic proximity to each other. These data hint at one of the several potential causative phenomena, particularly incomplete coalescence or gene flow. In other cases polymorphisms are shared among geographically more disparate groups of accessions; the genesis of these shared polymorphisms is unknown.

Origin of the Kimberley cottons

The three molecular data sets may be used in conjunction with the sequence divergence to infer the timing of past cladogenetic and biogeographic events. We take into account a number of caveats and limitations to the use of molecular clocks (Hillis et al. 1996), including the

possibility of substitution rate heterogeneity among lineages, uncertainty regarding clock calibrations, and unknown, but presumably large, estimation errors. Accordingly, we view these estimates as offering useful, albeit imprecise, approximations of divergence times.

Among the three data sets, ITS sequences were not used for divergence time estimation because these sequences exhibit significant rate heterogeneity within *Gossypium* (Wendel et al. 1995) and because rates of evolution for multigene families may be subject to more unknowns and larger variances than are those for single-copy genes. Before applying a clock based on *rpl16* and *AdhD*, we first evaluated whether substitution rates were homogeneous within Australian *Gossypium*. To accomplish this we employed the methods of Tajima (1993) using the program *Tajima93* (T. Seelanan, unpublished). Results for *rpl16* intron sequences indicated that nucleotide substitution rates are homogeneous, but this was not the case for *AdhD*. When the Tajima test was performed on the entire *AdhD* sequence or for introns only, significant substitution rate heterogeneity was found among many taxa; when, however, the test was performed using just exon data, only *G. sturtianum* exhibited a significantly ($P < 0.05$) higher substitution rate than other taxa. Collectively, these rate tests establish that substitution rates for *rpl16* intron and *AdhD* exons (excluding that of *G. sturtianum*) are homogeneous, and were therefore useful for divergence time estimations.

A critical aspect of divergence time estimation is clock calibration. For *rpl16*, substitution rates for tetraploid *Gossypium* have been estimated to be 0.4×10^{-9} to 1.1×10^{-9} substitutions per site per year (Wendel and Albert 1992; Seelanan et al. 1997; Small et al. 1998), whereas for *Gleditsia* and *Gymnocladus*, an intermediate rate has been published (0.6×10^{-9} substitutions per site per year; Schnabel and Wendel 1998). These rates, for non-coding intron sequences, logically are similar to the range of synonymous substitution rates reported (Eyre-Walker and Gaut 1997) for protein coding genes for palms (0.35×10^{-9}) and grasses (1.2×10^{-9}). For *AdhD*, we assumed previously published divergence times for *G. herbaceum* and *G. raimondii* (Wendel and Albert 1992; Seelanan et al. 1997) and used the *AdhD* sequence data for these same taxa (this study and Small, unpublished data) to calculate a rate of 3.6×10^{-9} to 7.2×10^{-9} substitutions

per site per year. These values approximate the synonymous substitution rates reported for *Adh* exon data from grasses and palms (Eyre-Walker and Gaut 1997). For each lineage of interest, the time since divergence was obtained by dividing mean Jukes-Cantor distances by twice the rate of nucleotide substitution. The mean Jukes-Cantor distance within each section was calculated from all pairwise comparisons among taxa within the respective section, excluding comparisons among taxa within sub-clades. For example, to calculate the mean Jukes-Cantor distance for *AdhD* exons within sect. *Grandicalyx*, we considered all Jukes-Cantor distances from pairwise comparisons, excluding pairwise comparisons among taxa in the clade of *G. exiguum-G. rotundifolium-G. pilosum-G. enthyle* (9239)-*G. marchantii* (499770)-*G. populifolium* (499808), and pairwise comparisons between taxa in the clade of *G. londonderriense* 9194-SC29 (see Tree **1** of Fig. 5). To calculate mean Jukes-Cantor distances among sections, all pairwise comparisons between taxa in one section and taxa in the other section were used. Because we used a range of rates for each calibration, as noted above, upper and lower bounds for divergence times were estimated.

The *rpl16* data were used to estimate divergences within the three taxonomic sections of Australian cottons. These calculations suggest that the two species in sect. *Sturtia* separated from one another about 3 to 8 million years before present (mybp), and that the earliest branching within sect. *Grandicalyx* occurred about 0.7 - 2 mybp. No variation was found between *G. australe* and *G. nelsonii*, and so divergence times within sect. *Hibiscoidea* could not be estimated. It should be noted that sequences of *G. bickii* and *G. cunninghamii* were not included in these estimations, due to their putatively introgressant histories. Using *AdhD* exon data, remarkably similar values were obtained. Specifically, species in sect. *Sturtia* are suggested to have diverged 3 - 6 mybp whereas species in sect. *Grandicalyx* diverged from each other starting about 1.7 - 3 mybp. The *AdhD* data also permitted estimation for species of sect. *Hibiscoidea*, which are estimated to have undergone cladogenesis 7 - 14 mybp.

We also used molecular clocks to estimate divergence among the three sections. Using *AdhD* data, the two arid zone sections, i.e., sect. *Sturtia* and sect. *Hibiscoidea*, are inferred to have separated from each other about 8 -

15 mybp. These calculations also suggested that section *Grandicalyx* departed from sect. *Sturtia* and sect. *Hibiscoidea* about 10.5 - 21 mybp and 11 – 22 mybp, respectively. This calculation highlights the long isolation of sect. *Grandicalyx* from the other two sections.

Because the maximum sequence divergence between any two species in sect. *Grandicalyx* is low, we suggest that extant Kimberley cottons evolved from a relatively recent common ancestor. The molecular clock calculations imply that this common ancestor is Pleistocene (*rpl16*) or late Pliocene/ Pleistocene (*AdhD*) in age. Although sect. *Grandicalyx* is indicated to have diverged from other Australian cottons at a much earlier age, perhaps in the Miocene, speciation among its extant members has been more recent. In this respect, the surviving members of sect. *Grandicalyx* may be about the same age as the tetraploid members of the genus (Wendel 1989; Small et al. 1998).

Historical phytogeography of sect. Grandicalyx

The foregoing divergence time estimates provide a historical framework for species in sect. *Grandicalyx*. These species have a geographic distribution that is restricted to the Kimberley Plateau in northwestern Western Australia, except for the single disjunct species, *G. cunninghamii*, from the Cobourg Peninsula of the Northern Territory. Although some species have overlapping geographic ranges, as depicted in our Fig. 1 and Figs. 1-5 of Fryxell et al. (1992), most species have distinct distributions and are almost always isolated from each other. Some species have highly restricted ranges and/or are known from few populations, such as *G. cunninghamii*, *G. marchantii*, and *G. pulchellum* whereas other species, such as *G. exiguum*, *G. rotundifolium*, and *G. pilosum* are more widespread and may even morphologically intergrade in regions of contact. Still other species have large aggregate ranges, but consist of only a few highly disjunct populations. For example, only six populations of *G. costulatum* are known, yet these few populations are separated by as much as 300 kilometers.

All three molecular phylogenetic hypotheses generated (Figs. 3-5) are remarkable for the absence of resolution among species of sect. *Grandicalyx*. This feature, when combined with the molecular clock results

and the distributional data, suggests that the phytogeographic history of the group may be as illustrated in Fig. 6. In this scenario, we postulate that the ancestor of modern species of sect. *Grandicalyx* had a wide geographic distribution obtained during more mesic times in the Miocene, and that it perhaps covered the entire Kimberley Plateau. A progressive increase in aridity resulted in range fragmentation of the ancestor into increasingly isolated geographic "pockets". Through time, this process led to local speciation and the present pattern of widely scattered localized populations.

This scenario is clearly speculative, but it is consistent with the phylogenetic pattern (unresolved "rakes") and divergence time estimates. It is also simplified, and modifications of the basic scenario are possible, such as early diversification into two lineages, one representing the prostrate taxa and the other those with a more upright habit. As an alternative scenario, one might postulate that present-day distribution reflects multiple long-distance dispersals over a short time period rather than range fragmentation of a widespread ancestor. We view this alternative as unlikely for several reasons, most notably because it unnecessarily invokes a mechanism (multiple long-distance dispersal events) which evidence suggests is unlikely. In particular, seeds of species in sect. *Grandicalyx* are apparently adapted for highly localized ant-dispersal (see below), and they contain terpenoid aldehydes (Brubaker et al. 1996) that may be a deterrent to dispersal by many other potential dispersal agents.

Palaeoclimatic and vegetation records of the Australian continent provide collateral support for the range-fragmentation scenario illustrated in Fig. 6. Specifically, the rainforest prevailing in northwestern Australia had disappeared by the Miocene, with tentative evidence suggesting that the palaeoclimate changed from mesic to more arid. It is thought that the climate continued to get drier towards the Pleistocene (Kershaw et al. 1994), perhaps as a consequence of the northward movement of Australia away from Antarctica (Beard 1977; Kemp 1978; Bowler 1982). As the rainforest shrank, a woody sclerophyllous forest gradually emerged, until the Pliocene, when, according to available evidence, a substantial increase in the herbaceous community occurred. Grasses in particular became prominent, leading to the development of a tropical savanna vegetation in the northwestern corner of the continent (Kershaw et al. 1994). This

palaeofloristic history coincides not only with the molecular data, but with morphological data as well. Specifically, species of sect. *Grandicalyx* are unique in the genus in being herbaceous perennials adapted to seasonal fires in grasslands. Thus, the climatic and ecological conditions favorable to the evolution of their present habit and habitat are congruent with the phylogenetic and phenetic data.

Evolutionary novelty in sect. Grandicalyx

Species in sect. *Grandicalyx* are distinguished from the remainder of the genus by several unusual morphological characteristics. Perhaps most obvious is the difference in growth form: whereas most species in *Gossypium* are upright shrubs to small trees, sect. *Grandicalyx* species are upright to decumbent to prostrate perennial herbs with an enlarged underground rootstock. All outgroups to sect. *Grandicalyx* are shrubby to arborescent, including most other genera in the tribe Gossypieae (Seelanan et al. 1997). It seems clear, therefore, that the herbaceous habit is derived. A plausible scenario for the origin of this habit (Fig. 7) is that it evolved in response to the increasing aridity and frequency of fires that occurred in the Kimberleys during the Pliocene and Pleistocene, as noted above. At present, the Kimberley region receives monsoon rains from November through March, and during the rest of the year there is usually no effective rainfall. Accordingly, fires are common during the dry season. Regeneration may be rapid, as all species in sect. *Grandicalyx* are capable of sending out new shoots from the thickened underground rootstocks following fires (CLB, JFW, JMS and LAC, pers. field observation; Craven et al. 1995).

In addition to the change in growth form, sect. *Grandicalyx* are novel in their seed-dispersal mechanism. With the single exception of one species endemic to Hawaii, *G. tomentosum* Nuttall ex Seemann, all species of *Gossypium* are associated with ants that feed on the nectar produced by the leaf and epicalyx nectaries. In general, it is not clear what benefit accrues to *Gossypium* species from nectar foraging by ants, but it may be that the ants ward off herbivory by other insects, as has been observed in cultivated cotton (Dhandapani et al. 1994; Fernandes et al. 1994) and *Hibiscus* spp. in the Kimberley area (JMS, per. field obs.). In sect. *Grandicalyx*, this association appears to have become more specialized, as all species have

become myrmecochorous, a trait found nowhere else in the genus. Myrmecochory was first introduced by Sernander (1906) to describe (as interpreted by Berg 1975) "... any plant whose diaspores (seeds, fruits, etc.) are sought out and disseminated by ants owing to the presence of particular ant-attracting tissue or structure which [Sernander] termed 'elaiosomes'." Seeds of sect. *Grandicalyx* species are black in color with sparse pubescence, but they are decorated with large (circa half or more the length of the seeds) white elaiosomes. These elaiosomes persist even after the seeds dry, although they shrivel to a fraction of the size evident in fresh seeds. Berg (1975) defined the typical characteristics of Australian myrmecochorous plants as being shrubs growing in xeric habitats with seeds possessing firm and durable elaiosomes. Species in sect. *Grandicalyx* occur in seasonally xeric environments, have the aforementioned elaiosomes, but are herbaceous perennials.

What is/are the selective advantage(s) of myrmecochory which stabilized this syndrome in the K-genome species? Beattie (1983) proposed five hypotheses, including: (1) the nutrient-rich nest environment promotes seedling establishment; (2) predator-avoidance through active dispersal to a protected site; (3) avoidance of competition for above-ground germination microsites; (4) fire-avoidance; and (5) dispersal over larger distances than if the dispersal mechanism was passive. The habitats where sect. *Grandicalyx* species are found include a diversity of floristic associations and soil types, but the most typical physiognomy is savanna grassland with an open, scattered *Eucalyptus* canopy. The fire-ecology of the region suggests one connection and potential selective explanation for the existence of ant-dispersal in the group, namely, fire-avoidance by the seeds. Another possible benefit of ant dispersal may be an increase in dispersal distance of seeds from parental plants, the fifth hypothesis of Beattie (1983). Support for this hypothesis is offered by the observation of ant species (*Rhytidoponera cerastes*) transporting intact seeds to their underground nests (Fryxell et al. 1992) and by our observation of active removal by an unidentified ant species of seeds from capsules immediately following dehiscence. Other ant species (*Chelaner* sp., *Monomorium* spp.) have been seen feeding on elaiosomes *in situ*, however, these species are too small to be able to transport the seeds. No study has been made of the fate of seeds

from species in sect. *Grandicalyx* after the elaiosomes are removed, so it is not known whether seeds remain in the ant nests or are removed from the nests as refuse. In the case reported by Fryxell et al. (1992) of *Rhytidoponera* mediated seed transfer, no *Gossypium* seeds were observed to have been deposited outside the nest (LAC, per. field obs.). We have observed situations where multiple plants arise from the same site as if there were multiple germinations from an underground nest, but it is not evident what the general pattern is. Accordingly, the first and third selective advantages proposed by Beattie (1983) need further investigation to ascertain whether these apply to Kimberley cottons. Finally, the hypothesis of predator-avoidance is difficult to reject at present. All sect. *Grandicalyx* species have seeds containing considerable amounts of terpenoid aldehydes (Brubaker et al. 1996) that have toxicity to a broad-spectrum of animals (Cherry and Leffler 1984). We have no information or direct observations on the degree to which insects, birds, or small mammals infest or harvest mature seeds of the Kimberley species. In some areas immature capsules were heavily predated by one or more Lepidopteran species, hence the proposed escape mechanism would not be valid for these. However, partially eaten mature seeds were found occasionally, indicating that birds or small mammals will consume the seed to some extent.

Future work

The present study addressed several aspects of the phylogeny and biology of *Gossypium* sect. *Grandicalyx*, yet many questions remain unanswered. From a cladistic standpoint, most of the history within the section remains to be elucidated. The apparent recent evolution of the group suggests that satisfactory resolution may be obtained only if many times more nucleotides are used than in the present study, and/or if other rapidly evolving characters are employed in phylogenetic analysis. These efforts will need to accommodate the necessity for wide sampling within most of the taxa, as recent evolution may coincide with the phylogenetically vexing problems caused by lineage sorting of polymorphisms or interspecific gene flow (Small et al. 1998; Wendel and Doyle 1998). These same phenomena, as well as the relative recency of the group, may also be

responsible for the some of the uncertainties that exist with respect to species delimitations and circumscriptions.

Additional phylogenetic work may also shed light on the evolutionary history of *G. cunninghamii*. Present data suggest strongly that it has a reticulate ancestry, but most of the details regarding this history remain unclear. This includes the timing of the introgression event(s), whether there has been partial nuclear introgression of markers that were not studied here, and whether this introgression is related to the geographic isolation of the species. The question arises as to whether novel or recombinant ecological requirements and/or capabilities emerged from the ancient hybridization.

Finally, detailed investigations of the mutualism between the Kimberley cottons and their ant associates may prove to be especially interesting. In addition to having relevance to questions of seed dispersal, species distribution, patterns of genetic variation, and the fate of seeds in the seed bank, nearly nothing is known about the species diversity of the ant associates nor their species-specificity. Such knowledge may shed light on the potential role of ants in plant defense, and may ultimately lead to information that has practical value in cotton production.

ACKNOWLEDGMENTS

We thank R. Small for assistance with methods for analysis of *Adh* genes and for outgroup sequences, R. Cronn for miscellaneous laboratory assistance, and E. D. Edwards for ant identification. Financial support was provided by The Institute for the Promotion of Teaching Science and Technology (Thailand), the National Geographic Society (USA.), the National Science Foundation (USA.), Cotton Research and Development Corporation (Australia), the US Department of Agriculture, and the Ben J. Altheimer Foundation.

LITERATURE CITED

BEARD, J. S. 1977. Tertiary evolution of the Australian flora in the light of latitudinal movements of the continent. Journal of Biogeography 4: 111-118.

BEATTIE, A. J. 1983. Distribution of ant-dispersed plants. Sonderbände des Naturwissenschaftlichen Vereins in Hamburg 7:249-270.

BENNETT, M. D., J. B. SMITH, and J. S. HESLOP-HARRISON. 1982. Nuclear DNA amounts in angiosperms. Proceedings of Royal Society of London series B 216: 179-199.

BERG, R. Y. 1975. Myrmecochorous plants in Australia and their dispersal by ants. Australian Journal of Botany 23: 475-508.

BOWLER, J. M. 1982. Aridity in the late Tertiary Quaternary of Australia. Pp. 35-45 in *Evolution of flora and fauna of arid Australia*, eds. W. R. Barker and P. J. M. Greenslade. Firewile: Peacock Publication.

BREMER, K. 1988. The limits of amino acid sequence data in agiosperm phylogenetic reconstruction. Evolution 42: 795-803.

BRUBAKER, C. L., C. G. BENSON, C. MILLER, and D. N. LEACH. 1996. Occurrence of terpenoid aldehydes and lysigenous cavities in the 'glandless' seeds of Australian *Gossypium* species. Australian Journal of Botany 44: 601-612.

CHERRY, J. P., and H. R. LEFFLER. 1984. Seed. In *Cotton, agronomy monograph* no. 24. R. J. Kohel and C. F. Lewis (eds.). pp. 511-569. Madison: Crop Science Society of America.

CRAVEN, L. A., J. MCD. STEWART, A. H. D. BROWN, and J. P. GRACE. 1995. The Australian wild species of *Gossypium*. In *Challenging the future: Proceedings of the world cotton research conference*. G. A. Constable and N. W. Forrester (eds.). pp. 178-181.

DHANDAPANI, N., S. JAYARAJ, and R.J. RABINDRA. 1994. Activity of ants on cotton plants sprayed with nuclear polyhedrosis virus and adjuvants against *Heliothis armigera* (Hubner). Journal of Entomological Research (New Delhi) 18: 65-68.

EYRE-WALKER A., and GAUT B. S. 1997. Correlated rates of synonymous site evolution across plant genomes. Molecular Biology And Evolution 14: 455-460.

ENDRIZZI, J. E., E. L. TURCOTTE, and R. J. KOHEL. 1985. Genetics, cytogenetics, and evolution of *Gossypium*. Advances in Genetics 23: 271-375.

FERNANDES, W.D., P.S. OLIVEIRA, S.L. CARVALHO, and M. E. M. HABIB. 1994. *Pheidole* ants as potential biological control agents of the boll weevil,

Anthonomus grandis (Col., Curculionidae), in Southeast Brazil. Journal of Applied Entomology 118: 437-441.

FRYXELL, P. A. 1965. A revision of the Australian species of *Gossypium* with observations on the occurrence of *Thespesia* in Australia (Malvaceae). Australian Journal of Botany 13: 71-102.

—. 1979. *The natural history of the cotton tribe.* College Station: Texas A & M University Press.

—. 1992. A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea 2: 108-165.

—, L. A. CRAVEN and J. McD. STEWART. 1992. A revision of *Gossypium* sect. *Grandicalyx* (Malvaceae), including the description of six new species. Systematic Botany 17: 91-114.

HILLIS, D. M., and J. P. HUELSENBECK. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. Journal of Heredity 83: 189-195.

—, B. K. MABLE, and C. MORITZ. 1996. Applications of molecular systematics. Pp. 515-543 in *Molecular systematics*, eds. D. M. Hillis, C. Moritz, and B. K. Mable. Sunderland: Sinauer.

JORDAN, W. C. M., M. W., COURNEY, and J. E. NEIGEL. 1996. Low level of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (Lemnaceae). American Journal of Botany 83: 430-439.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. In *Mammalian protein metabolism.* H. N. Munro (ed.). New York: Academic Press.

KELCHNER, S. A., and L. G. CLARK. 1997. Molecular evolution and phylogenetic utility of the chloroplast rpl16 intron in Chusquea and the Bambusoideae (Poaceae). Molecular Phylogenetics and Evolution 8: 385-397.

KELCHNER, S. A. and J. F. WENDEL. 1996. Hairpins create minute inversions in non-coding regions of chloroplast DNA. Current Genetics 30: 259-262.

KEMP, E. M. 1978. Tertiary climatic evolution and vegetation history in the southeast Indian Ocean region. Palaeogeography, Palaeoclimatography, Palaeoecology 24: 169-208.

KERSHAW, A. P., H. A. MARTIN, AND J. R. C. MCEWAN MASON. 1994. The neogene: A period of transition. Pp. 299-327 in *History of Australian vegetation: Cretaceous to recent*, ed. R. S. Hill. Cambridge: Cambridge University Press.

KUMAR, S., K. TAMURA, and M. NEI. 1993. MEGA: Molecular evolutionary genetic analysis, version 1.1. The Pennsylvania State University, University Park, PA.

MILLAR, A. A., and E. S. DENNIS. 1996. The alcohol dehydrogenase genes of cotton. Plant Molecular Biology 31: 897-904.

PALMER, J. D. 1991. Plastid chromosome: structure and evolution. Cell Culture and Somatics Cell Genetics of Plants 7A: 5-53.

PATERSON, A. H., C. L. BRUBAKER, and J. F. WENDEL. 1993. A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Molecular Biology Reporter 11: 122-127.

SAMBROOK, J., E. F. FRITSCH, AND T. MANIATIS. 1989. Molecular cloning: A laboratory manual 2nd ed. New York: Cold Spring Harbor Laboratory.

SCHNABEL, A., and J. F. WENDEL. 1998. Cladistic biogeography of *Gleditsia* (leguminosae) based on *ndhf* and *rpl16* chloroplast gene sequences. American Journal of Botany (in press).

SEELANAN, T., A. SCHNABEL, and J. F. WENDEL. 1997. Congruence and consensus in the cotton tribe (Malvaceae). Systematic Botany 22: 259-290.

SERNANDER, R. 1906. Monographie der Euroäischen myrmekochoren. Kungliga Svenska Vetenskapsakademiens Handlingar. 41 no. 7: 1-410.

SMALL, R. L., J. A. RYBURN, R. C. CRONN, T. SEELANAN, and J. F. WENDEL. 1998. The tortoise and the hare: Choosing between non-coding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. American Journal of Botany (in press).

STEWART, J. MCD. 1995. Potential for crop improvement with exotic germplasm and genetic engineering. Pp. 313-327 in *Challenging the Future: Proceedings of the World Cotton Research*, eds. G. A. Constable and N. W. Forrester (eds.). Conference-1. CSIRO, Melbourne.

STEWART, J. MCD., L. A. CRAVEN and J. F. WENDEL. 1997. A new Australian species of *Gossypium*. P. 448 in *Proceedings of Beltwide Cotton*

Conference. National Cotton Council, Memphis, TN.

SWOFFORD, D. L. 1993. PAUP: Phylogenetic analysis using parsimony, version 3.1. Champaign: Illinois Natural History Survey.

TAJIMA, F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135: 599-607.

WENDEL, J. F. 1989. New World tetraploid cottons contain Old World cytoplasm. Proceedings of the National Academy of Sciences USA 86: 4132-4136.

—, and V. A. ALBERT. 1992. Phylogenetics of the cotton genus (*Gossypium*): Character-state weighted parsimony analysis of chloroplast DNA restriction site data and its systematic and biogeographic implications. Systematic Botany 17: 115-143.

—, and J. J. DOYLE. 1998. Phylogenetic incongruence: Window into genome history and molecular evolution. In: Molecular Systematics of Plants II (P. Soltis, D. Soltis, and J. Doyle, eds.). Chapman and Hall (in press).

—, A. SCHNABEL, AND T. SEELANAN. 1995. An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. Molecular Phylogenetics and Evolution 4: 298-313.

—, J. McD. STEWART and J. H. RETTIG. 1991. Molecular evidence for homoploid reticulate evolution in Australian species of *Gossypium*. Evolution 45: 694-711.

TABLE 1. Taxa sampled. The taxonomic treatment follows Fryxell (1979, 1992), Stewart (1995), and Stewart et al. (1997). Accession designations are ours, except that those starting with PI are from the National Collection of *Gossypium* Germplasm (Percival 1987). [1] When available, coordinates in latitude and longitude are given. Otherwise, descriptive provenances are provided. WA= Western Australia; NT=Northern Territory. [2] Voucher specimens are deposited at Australian National Herbarium (*CANB*) and at Northern Territory Herbarium in Darwin (*DNA*) and at New York Botanical Garden Herbarium (NY); many duplicates are deposited at Ada Hayden Herbarium (*ISC*). The following abbreviations are used for collectors' names: AAM = A. A. Mitchell; AHDB = A.H.D. Brown; IC = Ian Cowie; CLB = C. L. Brubaker; EDE = E. D. Edwards; HH=H. Hurka; JW or JFW = J. F. Wendel; JMS = J. McD. Stewart; JPG= J. P. Grace; LAC = L. A. Craven; PAF= P. A. Fryxell; TDC=T.D. Couch; TH=T. Hymowitz; TSH = T.S. Henshaw.

Taxa/Accession	**Locality[1]**	**Voucher specimens[2]**
Ingroup taxa		
Section *Sturtia* ; Genomic designation: C		
Gossypium sturtianum J. H. Willis		
LC 9355	20°55'21"S 139°35'45"E	CLB & JFW 9355, *CANB*
PI 499810	Namatjira Dr. 22 km E of Glen Helen, NT	
Gos 5076	30°15'S 149°59'E	LAC 8822, *CANB*
Gos 5119	30°11'S 139°26'E	JPG, AHDB, LAC, & HH 1078, *CANB*
Gos 5237	23°1S 119°10'E	JPG, AHDB, & TH 1361, *CANB*

TABLE 1 (continued).

Gossypium robinsonii F. Mueller		
LC 9297	20°40'45"S 117°08'48"E	LAC & JMS 9297, *CANB*
LC 9300	21°34'37"S 117°03'27"E	LAC & JMS 9300, *CANB*
Section *Hibiscoidea* Tod. ; Genomic designation: G		
Gossypium bickii Prokh.		
LC 9360	20°20'S 139°10'E	CLB, & JFW 9360, *CANB*
PI 464843	19°18'S 129°36'E	TSH 2309, *DNA*
Gos 5048	23°44'S 133°56'E	PAF, LAC, & JMS 4463, *CANB*
Gossypium nelsonii Fryxell		
LC 9356	20 Km S of Mt. Isa, Ql, along Dajarra Road (Rt 83)	CLB & JFW 9356, *CANB*
PI 499783	Ormiston Gorge, NT	
Gossypium australe F. Mueller		
LC 9361	75.5 Km NW of Mt Isa, Ql, along Barkly Hwy. (Rt 66).	CLB & JFW 9361, *CANB*
PI 478754	Carawine Gorge, E. of Marble Bar, WA	
PI 499756	Ormiston Gorge, NT	PAF, LAC, & JMS 4460, *CANB*
Section Grandicalyx (Fryxell) Fryxell ; Genomic designation: K		
Gossypium anapoides J. M. Stewart, Craven & J. F. Wendel		

TABLE 1 (continued).

LC 9188	13°54'S 126°48'E	LAC, JMS, & JFW 9188, *CANB*
LC 9192	13°48'08"S 126°45'01"E	LAC, JMS, & JFW 9192, *CANB*
LC 9193	13°46'15"S 126°49'03"E	LAC, JMS, & JFW 9193, *CANB*
Gossypium costulatum Tod.		
LC 9241	15°31'53"S 124°37'04"E	LAC, JMS, & JFW 9241, *CANB*
LC 9260	15°02'17"S 125°02'45"E	JMS, CLB, & EDE 9260, *CANB*
LC 9262	15°15'20"S 125°06'07"E	JMS, CLB, & EDE 9262, *CANB*
Gossypium cunninghamii Tod.		
PI 499775	11°13'S 132°05'E	PAF, LAC, & JMS 4926, *CANB*
PI 499778	11°21'S 132°32'E	PAF, LAC, & JMS 4932 *CANB*
Gossypium enthyle Fryxell, Craven & J. M. Stewart		
LC 9224	15°13'39"S 125°41'01"E	LAC, JMS, & CLB 9224, *CANB*
LC 9238	15°10'38"S 125°42'05"E	LAC, JMS,& JFW 9238, *CANB*
LC 9239	15°00'01"S 125°40'21"E	LAC, JMS,& JFW 9239, *CANB*
Gossypium exiguum Fryxell, Craven & J. M. Stewart		
IC 4223	15°09'22"S 125°43'53"E	CI 4223 & JMS, *DNA*
LC 9151	15°30'S 126°20'E	LAC, JMS,JFW,&CLB 9151, *CANB*
LC 9225b	15°05'38"S 125°48'21"E	LAC & CLB 9225, *CANB*
LC 9235	14°50'S 125°43'56"E	LAC, JMS, & JFW 9235, *CANB*
Gos 5182	15°41'S 125°43E	IC 4309 & JMS, *DNA*

TABLE 1 (continued).

Gossypium londonderriense Fryxell, Craven & J. M. Stewart		
LC 9194	13°46'57"S 126°57'19"E	LAC, JMS, & JFW 9194, *CANB*
LC 9200	14°04'21"S 127°08'45"E	LAC, JMS, & JFW 9200, *CANB*
SC-29	14°06'21"S 127°09'40"E	LAC & JMS 9335, *CANB*
Gossypium marchantii Fryxell, Craven & J. M. Stewart		
IC 4203	14°02'14"S 125°59'14"E	CI 4203 & JMS, *DNA*
IC 4211	14°03'34"S 126°04'33"E	CI 4211 & JMS, *DNA*
IC 4219	14°06'54"S 126°09'28"E	CI 4219 & JMS, *DNA*
PI 499770	14°11'S 126°10'E	JMS, *ISC*
Gossypium nobile Fryxell, Craven & J. M. Stewart		
IC 4229	14°18'02"S 126°39'05"E	CI 4229 & JW, *DNA*
IC 4231	14°27'01"S 126°40'16"E	CI 4231, *DNA*
IC 4235	14°29'10"S 126°43'19"E	CI 4235 & JW, *DNA*
Gossypium pilosum Fryxell		
LC 9252	14°27'32"S 125°51'02"E	JMS, JFW, & EDE 9252, *CANB*
PI 478760	Lone Dingo on Mitchell Plateau, WA	PAF, LAC, & JMS 4039, *CANB*
PI 499785	14°33'S 125°51'E	
PI 499786	Mitchell Plateau, 16 km N of AMEX Camp, WA	JMS *s.n.*, *ISC*

TABLE 1 (continued).

PI 499803	14°31'S 125°48'E	PAF, LAC, & JMS 4556, *CANB*
CampRd	Camp Road (approx., 3 km W of AMEX Camp), WA	
Gossypium populifolium (Benth.) F. Mueller ex Tod.		
IC 4257	14°17'31"S 126°34'14"E	IC & CLB 4257, *DNA*
IC 4278	14°16'36"S 126°32'46"E	CI 4278 & CLB, *DNA*
LC 9203	15°20'36"S 124°33'47"E	LAC, JMS, & CLB 9203, *CANB*
LC 9209	15°17'S 124°31'E	LAC, JMS, & CLB 9209, *CANB*
LC 9250	14°23'17"S 126°23'59"E	JMS, JFW, & EDE 9250, *CANB*
PI 499808	14°50'S 126°30'E	PAF, LAC, & JMS 4856, *CANB*
Gossypium pulchellum (C. A. Gardner) Fryxell		
LC 9169	14°14'S 126°13'E	LAC, JMS, & CLB 9169, *CANB*
PI 464858	S of Jar Island, Vansittart Bay, WA	JMS 20, *NY*
Gos 5204	14°14'S 126°13'E	LAC, JMS, & CLB, 9177, *CANB*
Gossypium rotundifolium Fryxell, Craven & J. M. Stewart		
LC 9270	16°00'17"S 125°55'04"E	JFW & CLB 9270, *CANB*
LC 9278	17°31'51"S 122°16'05"E	LAC, JMS, & AAB 9278, *CANB*
PI 499787	17°45'S 122°15'E	PAF, LAC, & JMS 4556, *CANB*
PI 499789	16°55'S 126°30'E	PAF, LAC, & JMS 4628, *CANB*

TABLE 1 (continued).

Gos 5212	16°6'S 126°31'E	IC 4362, *DNA*
Gos 5217	17°30'E 122°25E	LAC, JMS, & AAM, 9281; *CANB*
Outgroup taxon		
Gossypium arboreum L.		
A_2-74		JFW & TDC 305
Gossypium herbaceum L.		
A_1-73		JFW 539

TABLE 2. Characteristics of *AdhD*. [1] Calculated as mean percentage difference of all pairwise comparisons among species (gaps excluded from calculation). Abbreviations are as following: sect. S = sect. *Sturtia*; sect. H = sect. *Hibiscoidea*; sect. G = sect. *Grandicalyx*. [2] Mean difference among species, calculated using Jukes and Cantor (1969) correction for multiple hits (gaps excluded). See Table 1 for taxa in each genomic group. For nucleotide substitution in exons, three values are given: nucleotide substitution per site; synonymous substitution per synonymous site (Ks); and nonsynonymous substitution per nonsynonymous site (Ka). [3] Total number of variable (and potentially phylogenetically informative) sites among sequences from ingroup taxa.

			Divergence [1]				Nucleotide substitution per site [2]				
Region	Aligned length (bp)	%GC	IG/ OG	sect. *S*	sect. *H*	sect. *G*	IG/ OG	sect. *S*	sect. *H*	sect. *G*	Substitution [3]
Exon	684	48.9	1.32	0.58	1.04	0.19	0.0133	0.0059	0.0105	0.0019	33 (13)
							0.0446 ^{Ks}	0.0190 ^{Ks}	0.0255 ^{Ks}	0.0075 ^{Ks}	
							0.0062 ^{Ka}	0.0029 ^{Ka}	0.0077 ^{Ka}	0.0004 ^{Ka}	
Intron	917	33.4	2.43	1.41	1.41	0.17	0.0247	0.0142	0.0143	0.0017	81 (28)
Overall	1601	40.2	1.82	0.96	1.20	0.18	0.0184	0.0096	0.0122	0.0018	114 (41)

FIGURE LEGENDS

FIG. 1. Collection sites of *Gossypium* species from Australia used in this study. Inset shows detail for the Kimberley region. Collections were made on expeditions conducted in 1981, 1983, 1985 and 1993.

FIG. 2. **Top**. The region from exon 2 to exon 8 of *AdhD*. Black blocks with arabic numerals are exons and exon numbers, respectively; white blocks with roman numerals are introns and intron numbers, respectively. Arrows pointing to the left are reverse primers whereas arrows pointing to the right are forward primers. **Bottom**. Autoradiograph of Southern hybridization experiments for copy number estimation for *AdhD* in representative accessions. In all accessions, only a single band of *AdhD* was detected for both enzymes, suggesting that *AdhD* is a single-copy gene in all taxa studied.

FIG. 3. Phylogenetic trees estimated from the *rpl16* intron data set. Tree **1** is from parsimony analysis when gaps were treated as missing data or were removed from the data matrix. The same tree was found when gaps were treated as presence/absence characters, in combination with their treatment either as missing data or when removed from data matrix, except that the topology in the dashed box was replaced with the four topologies (**A** through **D**) shown in the inset. Branch lengths are proportional to the number of characters that changed. Numbers above and below branches are the minimum number of characters supporting each branch and Bremer support, respectively. Tree **2** is a neighbor-joining tree based on analysis of Jukes-Cantor distances.

FIG. 4. Phylogenetic trees estimated from the ITS data set. Tree **1** is the strict consensus recovered in parsimony analysis. Numbers above and below branches are the minimum number of characters supporting each branch and Bremer support, respectively. Tree **2** is a neighbor-joining tree based on Jukes-Cantor distances.

FIG. 5. Phylogenetic trees estimated from the *AdhD* data set. Tree **1** is the strict consensus recovered in parsimony analysis estimated from the combined exon + intron data. When the analysis was based on exon data only, the strict consensus tree obtained is similar to Tree **1** except that clades shown by striped lines are not present. If only intron data are used,

the strict consensus tree is similar to Tree **1** except that branches represented by thick lines are not present. Numbers above and below branches are the minimum number of characters supporting each branch and Bremer support, respectively. Tree **2** is a neighbor-joining tree based on Jukes-Cantor distances. Circled **X** and **I** symbols to the right of Tree **2** denote branches retained in NJ trees inferred from only exon and only intron regions, respectively.

FIG. 6. A scenario for the historical phytogeography of *Gossypium* sect. *Grandicalyx* from NW Australia. For simplicity, only one widely distributed ancestral species is indicated, although several other possibilities exist, such as the occurrence of several widespread ancestral taxa. See text for details.

FIG. 7. An hypothesis for morphological and ecological evolution in Australian *Gossypium*. Shown at the root of the tree is the ancestral condition that characterizes all other *Gossypium*, viz., shrubs to small trees to subshrubs in seasonally arid to arid habitats, with seeds variously pubescent but almost always with a well-defined layer of epidermal hairs, with extrafloral nectaries (leaf and epicalyx) that are often visited by ants, and with genomes that were, until recently, thought to be the largest in the genus (circa 5 pg/2C, Bennett et al. 1982). Morphological transformations accompanying the recent radiation in the Kimberley region of NW Australia include a change in habit to either upright or prostrate perennial herbs, adaptation to a strongly monsoon climate characterized by extreme seasonal aridity, loss of most seed pubescence and development of a prominent elaiosome, a more intimate association with ants, and an increase in genome size to the largest in the genus (circa 7 pg/2C; Stewart, 1995).

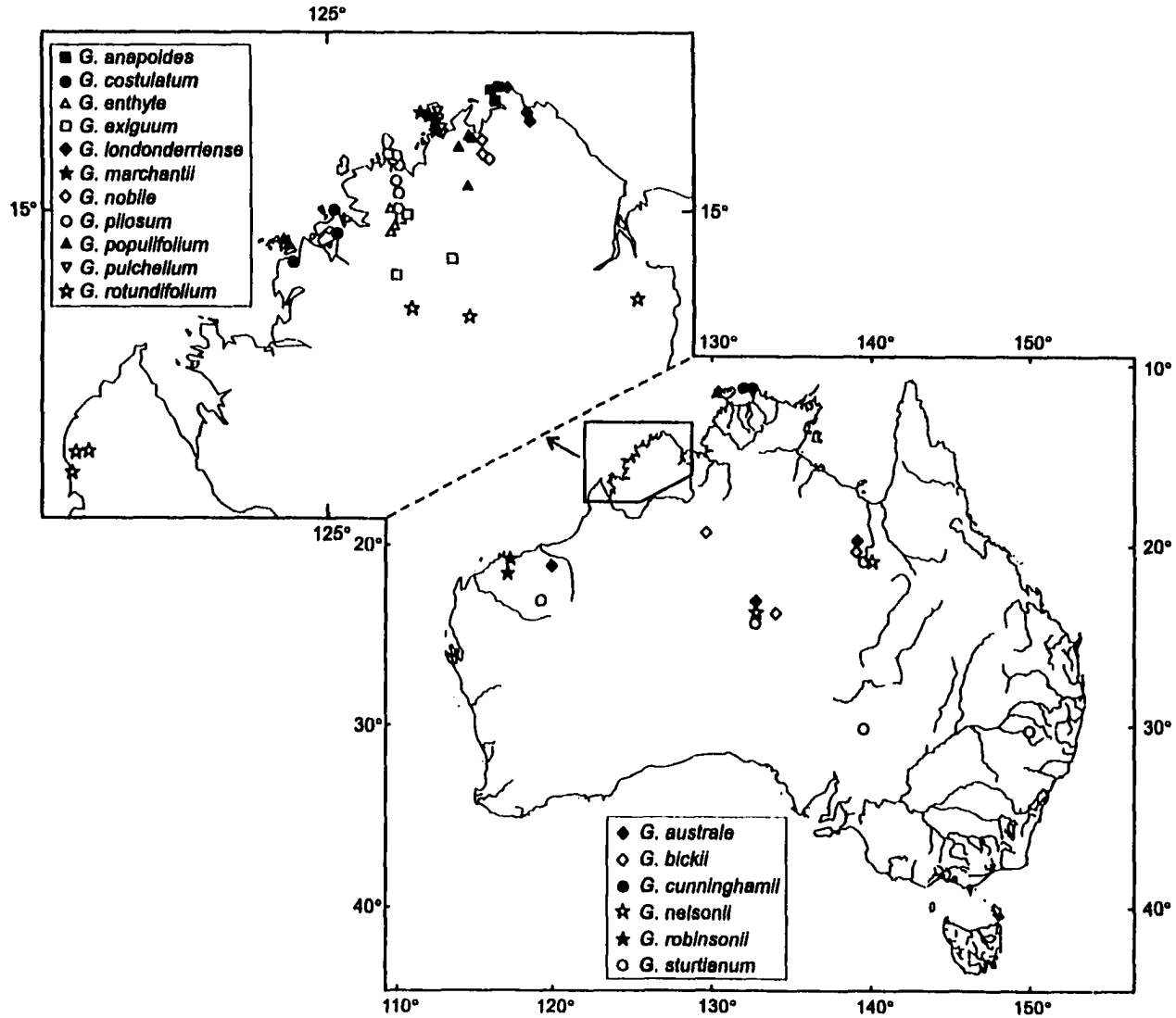

G. anapoides
G. costulatum
G. enthyte
G. exiguum
G. londonderriense
G. marchantii
G. nobile
G. pilosum
G. populifolium
G. pulchellum
G. rotundifolium
G. australe
G. bickii
G. cunninghamii
G. nelsonii
G. robinsonii
G. sturtianum
125°
15°
130°
140°
150°
10°
20°
30°
40°
110°
120°

Fig. 1

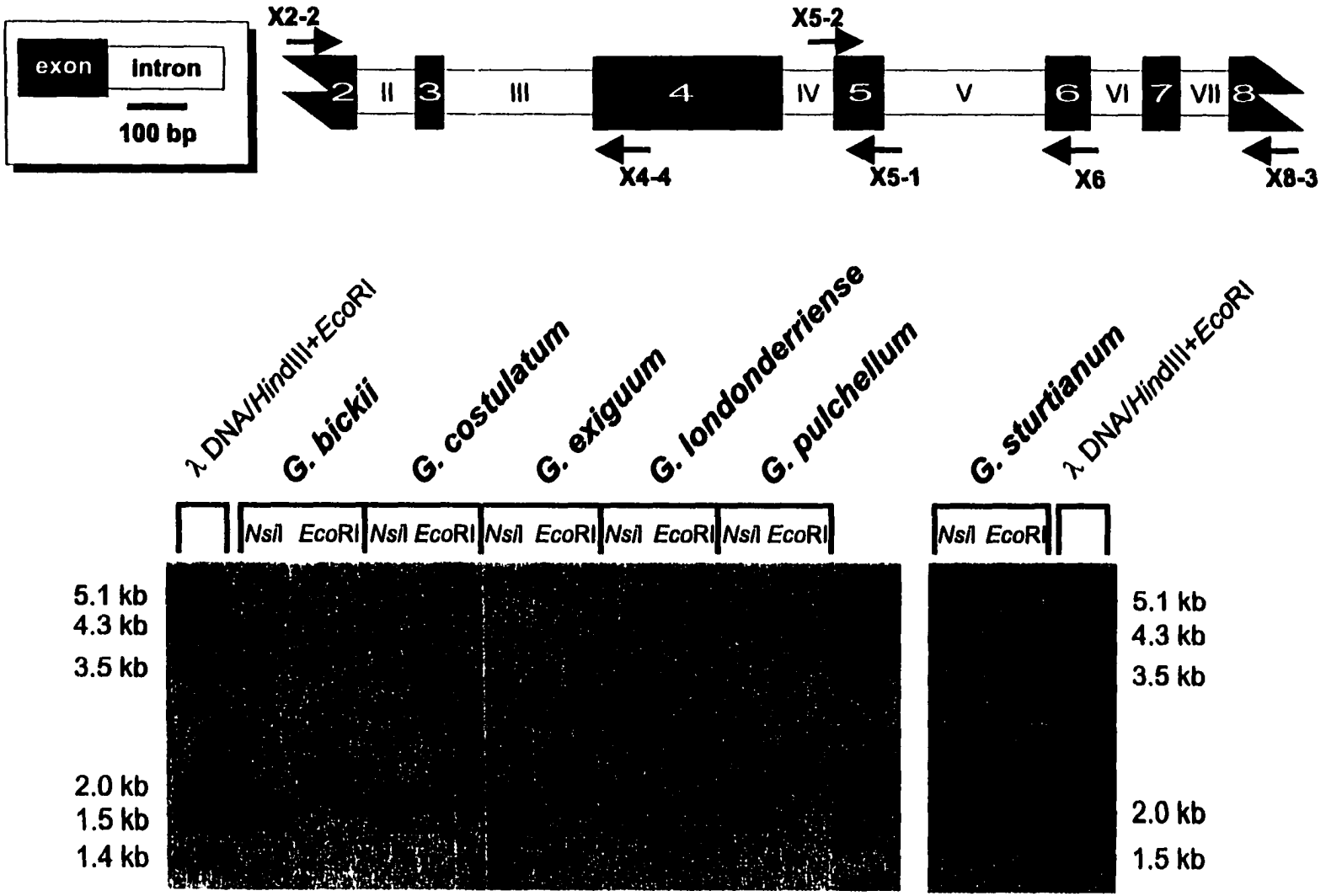

exon
Intron
100 bp
X2-2
X5-2
2
II
3
III
4
IV
5
V
6
VI
7
VII
8
X4-4
X5-1
X6
X8-3
λ DNA/HindIII+EcoRI
G. bickii
G. costulatum
G. exiguum
G. londonderriense
G. pulchellum
G. sturtianum
λ DNA/HindIII+EcoRI
NsiI EcoRI
NsiI EcoRI
NsiI EcoRI
NsiI EcoRI
NsiI EcoRI
NsiI EcoRI
5.1 kb
4.3 kb
3.5 kb
2.0 kb
1.5 kb
1.4 kb
5.1 kb
4.3 kb
3.5 kb
2.0 kb
1.5 kb

Fig. 2

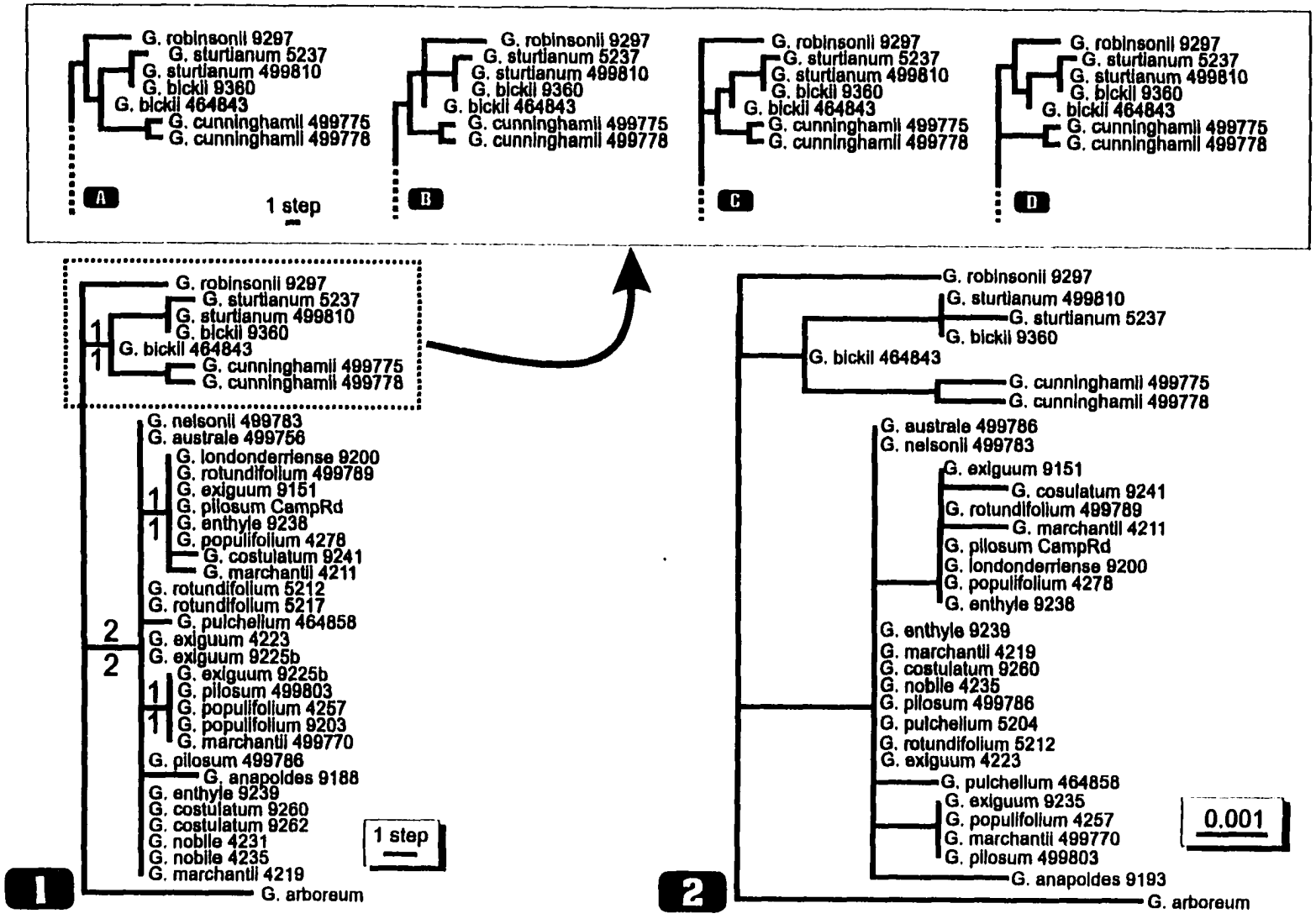

G. robinsonii 9297
G. sturtianum 5237
G. sturtianum 499810
G. bickii 9360
G. bickii 464843
G. cunninghamii 499775
G. cunninghamii 499778
A
1 step
B
C
D
1
2
G. nelsonii 499783
G. australe 499756
G. londonderriense 9200
G. rotundifolium 499789
G. exiguum 9151
G. pilosum CampRd
G. enthyle 9238
G. populifolium 4278
G. costulatum 9241
G. marchantii 4211
G. rotundifolium 5212
G. rotundifolium 5217
G. pulchellum 464858
G. exiguum 4223
G. exiguum 9225b
G. pilosum 499803
G. populifolium 4257
G. populifolium 9203
G. marchantii 499770
G. pilosum 499786
G. anapoides 9188
G. enthyle 9239
G. costulatum 9260
G. costulatum 9262
G. nobile 4231
G. nobile 4235
G. marchantii 4219
G. arboreum
1 step
G. australe 499786
G. nelsonii 499783
G. exiguum 9151
G. cosulatum 9241
G. marchantii 4211
G. enthyle 9239
G. marchantii 4219
G. costulatum 9260
G. nobile 4235
G. pilosum 499786
G. pulchellum 5204
G. exiguum 9235
G. anapoides 9193
0.001

Fig. 3

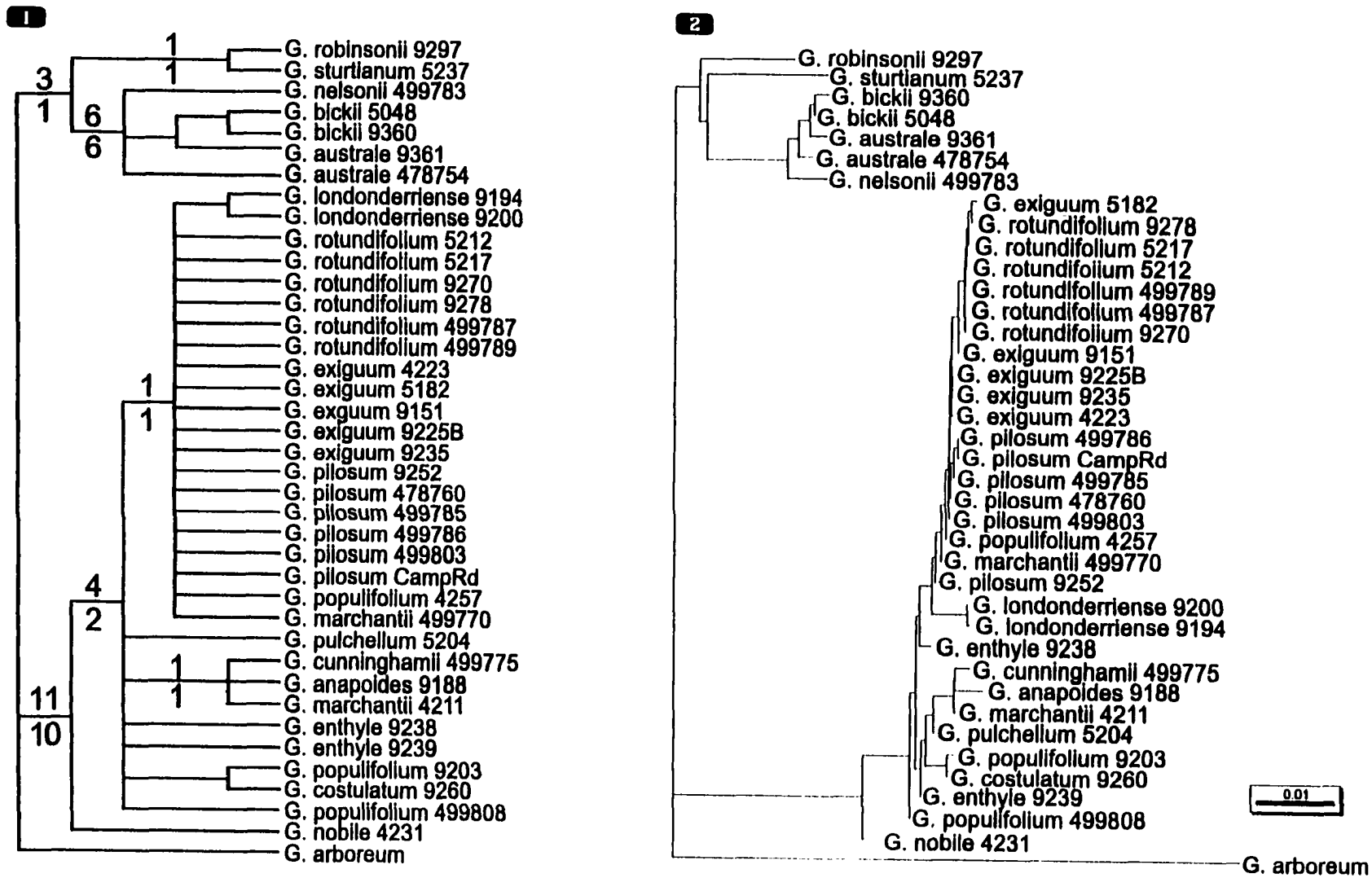

1
G. robinsonii 9297
G. sturtianum 5237
G. nelsonii 499783
G. bickii 5048
G. bickii 9360
G. australe 9361
G. australe 478754
G. londonderriense 9194
G. londonderriense 9200
G. rotundifolium 5212
G. rotundifolium 5217
G. rotundifolium 9270
G. rotundifolium 9278
G. rotundifolium 499787
G. rotundifolium 499789
G. exiguum 4223
G. exiguum 5182
G. exguum 9151
G. exiguum 9225B
G. exiguum 9235
G. pilosum 9252
G. pilosum 478760
G. pilosum 499785
G. pilosum 499786
G. pilosum 499803
G. pilosum CampRd
G. populifolium 4257
G. marchantii 499770
G. pulchellum 5204
G. cunninghamii 499775
G. anapoides 9188
G. marchantii 4211
G. enthyle 9238
G. enthyle 9239
G. populifolium 9203
G. costulatum 9260
G. populifolium 499808
G. nobile 4231
G. arboreum
2
G. robinsonii 9297
G. sturtianum 5237
G. bickii 9360
G. bickii 5048
G. australe 9361
G. australe 478754
G. nelsonii 499783
G. exiguum 5182
G. rotundifolium 9278
G. rotundifolium 5217
G. rotundifolium 5212
G. rotundifolium 499789
G. rotundifolium 499787
G. rotundifolium 9270
G. exiguum 9151
G. exiguum 9225B
G. exiguum 9235
G. exiguum 4223
G. pilosum 499786
G. pilosum CampRd
G. pilosum 499785
G. pilosum 478760
G. pilosum 499803
G. populifolium 4257
G. marchantii 499770
G. pilosum 9252
G. londonderriense 9200
G. londonderriense 9194
G. enthyle 9238
G. cunninghamii 499775
G. anapoides 9188
G. marchantii 4211
G. pulchellum 5204
G. populifolium 9203
G. costulatum 9260
G. enthyle 9239
G. populifolium 499808
G. nobile 4231
G. arboreum
0.01

Fig. 4


1
2
2
10
10
1
1
1
1
14
13
1
1
G. robinsonii 9297
G. sturtianum 5237
G. nelsonii 499783
G. australe 9361
G. bickii 9360
G. londonderriense 9194
G. londonderriense SC-29
G. londonderriense 9200
G. costulatum 9262
G. rotundifolium 5212
G. rotundifolium 5217
G. rotundifolium 9270
G. rotundifolium 9278
G. rotundifolium 499789
G. rotundifolium 499787
G. exiguum 4223
G. exiguum 5182
G. exiguum 9151
G. exiguum 9225B
G. exiguum 9235
G. pilosum 9252
G. pilosum 478760
G. pilosum 499785
G. pilosum 499786
G. pilosum 499803
G. pilosum CampRd
G. enthyle 9239
G. populifolium 499808
G. marchantii 499770
G. pulchellum 5204
G. pulchellum 464858
G. cunninghamii 499775
G. anapoides 9188
G. enthyle 9224
G. populifolium 4257
G. populifolium 4278
G. populifolium 9203
G. populifolium 9209
G. populifolium 9250
G. costulatum 9260
G. costulatum 9241
G. nobile 4229
G. nobile 4231
G. nobile 4235
G. marchantii 4203
G. marchantii 4211
G. marchantii 4219
G. herbaceum
2
G. robinsonii 9297
G. sturtianum 5237
G. nelsonii 499783
G. australe 9261
G. bickii 9360
G. costulatum 9241
G. populifolium 9203
G. populifolium 9209
G. pulchellum 5204
G. cunninghamii 499775
G. nobile 4229
G. nobile 4231
G. nobile 4235
G. londonderriense 9194
G. londonderriense SC-29
G. londonderriense 9200
G. costulatum 9262
G. anapoides 9188
G. costulatum 9260
G. pulchellum 464858
G. marchantii 4203
G. marchantii 4211
G. marchantii 4219
G. enthyle 9224
G. populifolium 4257
G. populifolium 9250
G. populifolium 4278
G. pilosum 478760
G. exiguum 5182
G. marchantii 499770
G. rotundifolium 9270
G. exiguum 4223
G. exiguum 9225B
G. rotundifolium 9278
G. rotundifolium 499787
G. rotundifolium 5212
G. rotundifolium 499789
G. rotundifolium 5217
G. pilosum 499785
G. exiguum 9151
G. pilosum 499803
G. populifolium 499808
G. pilosum CampRd
G. pilosum 499786
G. exiguum 9235
G. enthyle 9239
G. pilosum 9252
G. herbaceum
0.01
X
I/X
I/X
I/X
I/X
I
I
X/I

Fig. 5

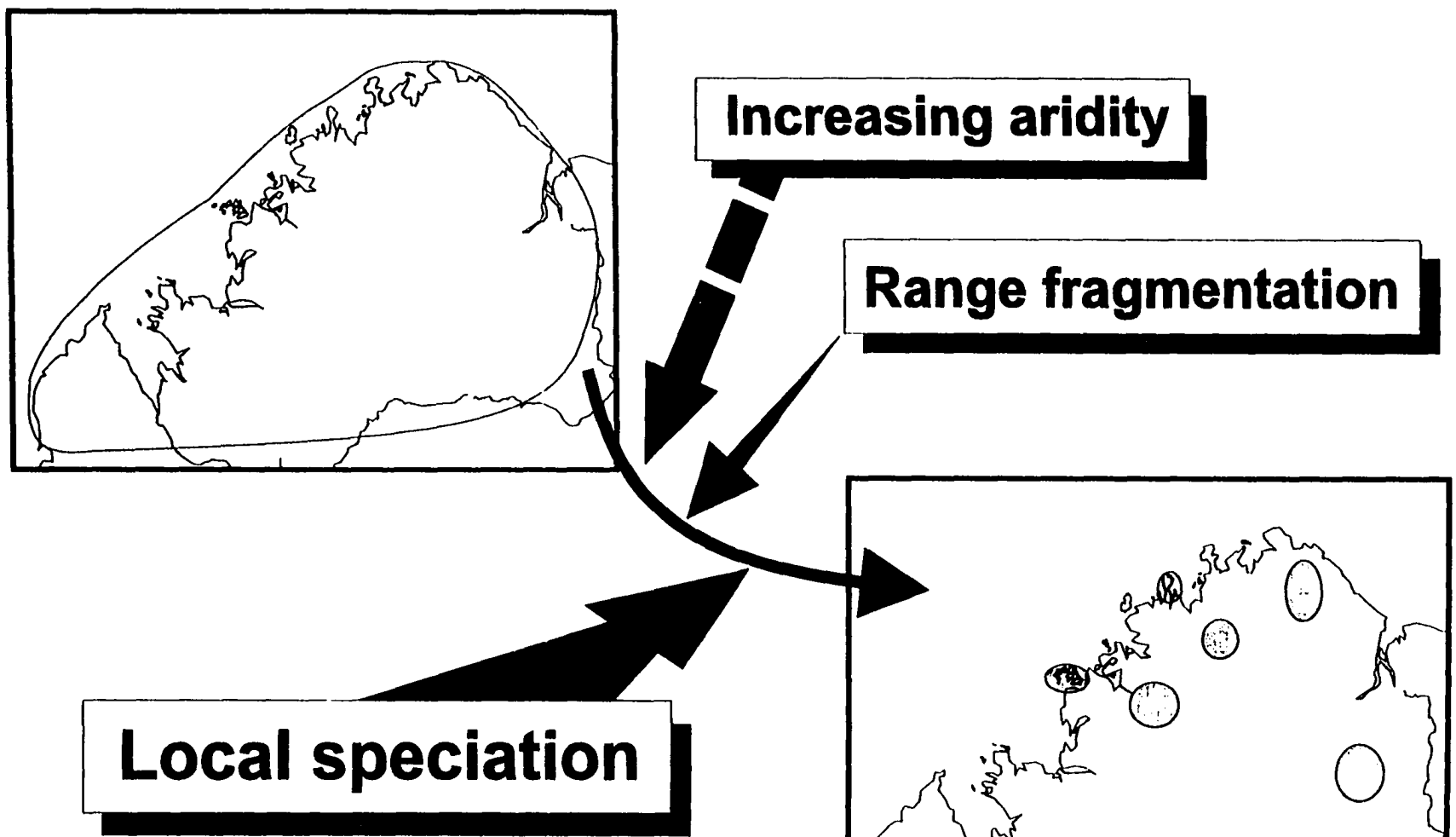

Increasing aridity
Range fragmentation
Local speciation

Fig. 6

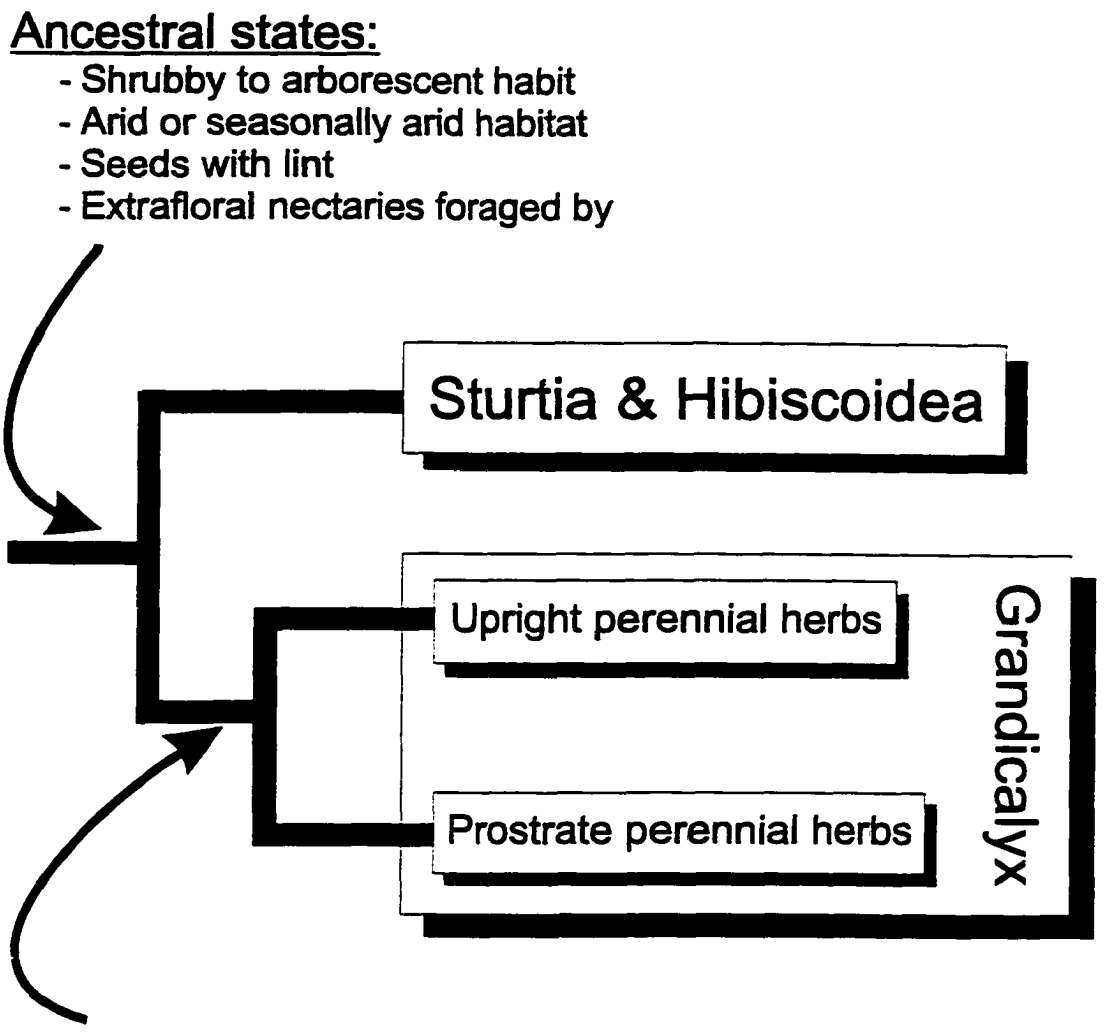

Ancestral states:
- Shrubby to arborescent habit
- Arid or seasonally arid habitat
- Seeds with lint
- Extrafloral nectaries foraged by
Sturtia & Hibiscoidea
Upright perennial herbs
Prostrate perennial herbs
Grandicalyx
Derived states:
- Upright to prostrate habit
- Adaptation to monsoon climate and seasonal fire
- Seeds without lints but bearing elaiosomes
- Development of myrmecochory
- Genome size 7 pg/2C

Fig. 7

CHAPTER 4. GENERAL CONCLUSIONS

This dissertation describes efforts to address questions of the phylogeny of the tribe Gossypieae, the genus *Gossypium*, and specifically *Gossypium* sect. *Grandicalyx*, using multiple molecular data sets. A strategy to handle multiple data sets was developed, which has been applied at the tribal and generic levels, but, however, could not be applied at the sectional level. This latter point is noteworthy because it may indicate that multiple data sets should be dealt with differently in phylogenetic studies depending on the taxonomic level of analysis. At tribal and generic levels, phylogenies inferred from each data set have more structure, i.e., the resulting tree is fully or nearly resolved, which facilitates further character incongruence tests. However, at the sectional level, phylogenetic reconstruction yields phylogenies with little to no structure, i.e., trees with little to no resolution, especially reconstructions using the *rpl16* intron data set. It is advisable not to perform character incongruence tests when level of resolution in different phylogenetic trees differs because the tests may give misleading results (Mason-Gamer and Kellogg 1996). In addition, parallel taxon sampling in molecular data sets at the sectional level could not be made due to intraspecific variation found in all three data sets. Accordingly, no further character incongruence tests were performed at the sectional level. Instead, a phylogeny of *Gossypium* sect. *Grandicalyx* was implicitly inferred from consensus trees generated from individual data sets.

Molecular data, particularly DNA sequences, have been regarded as one of the best tools for phylogenetic studies as they provide far more characters and character states than morphological characters. However, in this study, despite the large amount of molecular data generated, several phylogenetic resolutions could not be reached. These uncertainties include (i) relationships among genera (*Hampea*, *Lebronnecia*, *Thespesia* sect. *Lampas*), (ii) relationships among the groupings (*Hampea*, *Lebronnecia*, *Thespesia* sect. *Lampas*), (*Gossypium*, *Kokia*, *Gossypioides*), and *Thespesia* sect. *Thespesia*, (iii) relationships among subgenera and sections within *Gossypium*, and (iv) relationships among species within *Gossypium* sect. *Grandicalyx*. In addition, there were biases with respect to the distribution of characters supporting phylogenetic inferences. At the tribal level,

although it was indicated that the lineage represented by *Cienfuegosia* is the basal lineage in the tribe Gossypieae, only one character supported this inference, a two-amino-acid deletion in *ndhF* that was found in all members of the tribe but *Cienfuegosia* and the outgroups. However, with the same data set the relationship between *Kokia* and *Gossypioides* as sister taxa and the monophyly of *Gossypium* were supported by as many as 6 characters each.

At the generic level in *Gossypium*, the clades consisting of species from the same genomic designations received higher character support than the clades indicating relationships among genomic designations. The bias in character support was more exaggerated in the phylogenetic study of Australian *Gossypium* as most characters went to support the monophyly of each section with few characters left for inferring phylogenetic relationships among species within each section. Nonetheless, there is one important implication regarding the bias in character distribution over phylogenetic tree(s). That is, either rapid or recent speciation in evolution of the tribe or genus or subgenus might be suggested when a corresponding clade has few character supports compared to an immediately preceding or succeeding clade. One obvious case is the early diversification into three sections (*Sturtia*, *Hibiscoidea*, and *Grandicalyx*) in evolution of Australian *Gossypium*, followed by the recent speciation in sect. *Grandicalyx*.

One advantage of molecular data over morphological characters is that estimates of divergence time can be obtained from molecular data. Although times estimated from molecular data are uncertain and subject to several sources of error, they provide an approximate time scale, which, when coupled with other evidence (e.g., palynological, palaeogeographical data), may contribute to an understanding of the history. This process was applied to divergence in the tribe Gossypieae and in *Gossypium*. Divergence time estimates suggested that the tribe Gossypieae originated during the Miocene, and that diversification among genera occurred in the mid-Miocene to mid-Pliocene, with the exception of the *Kokia-Gossypioides* separation. These two genera diverged 3 million years ago, or in the late Pliocene, according to the molecular data. The implication from this estimate, when considered in conjunction with their geographic distribution, implicates transoceanic dispersal as a factor in the evolution of these two

genera. This finding reinforces long-distance, saltwater dispersal as the major mechanism in seed dispersal in the tribe Gossypieae. In *Gossypium*, estimates of divergence time indicated the major genomic groups, viz., A, C, and D genomes, diverged in a temporally closely spaced manner in the early evolution of the genus. Molecular clock estimates also suggested recent (Pliocene-Pleistocene) speciation in Australian *Gossypium* sect. *Grandicalyx*.

This dissertation also suggests trends in the evolution of morphological characters in the tribe Gossypieae and in *Gossypium*. Most members of the Gossypieae and *Gossypium* are shrubby or sub-arborescent woody plants. In *Cienfuegosia* and in Australian *Gossypium* sect. *Grandicalyx*, however, they are either small shrubs or decumbent to prostrate herbaceous perennials. In addition, the usual condition in the tribe and the genus is that species have seed bearing dense lints; exceptions include species in *Hampea* and in Australian *Gossypium* sect. *Grandicalyx*, whose seeds bear sparse to no lint, and instead, have prominent elaiosomes. When considered in a phylogenetic context, these two groups are derived. It is reasonable, therefore, to postulate that the transition in morphological characters is from the woody plants with an arborescent habit with densely linted seeds, to herbaceous plants with a decumbent habit with sparsely pubescent to lintless seeds that bear elaiosomes.

A related character transformation indicated by the present work concerns mechanism of seed dispersal. As previous mentioned, long-distance, saltwater dispersal has been a major factor in the evolution of the tribe and of *Gossypium*. In Australian *Gossypium* sect. *Grandicalyx*, however, seed dispersal by ants prevails. Furthermore, it has been found that other species of Australian *Gossypium*, particularly *G. australe*, have seeds that can be carried by wind. It is likely that the transition in seed dispersal is from passive mechanisms, with occasional opportunistic oceanic voyages, to a more recent wind- and ant-mediated mechanism.

Although several phylogenetic inferences were advanced in this dissertation, there remain many unanswered questions. The following are suggested further studies:

(i) In future phylogenetic studies of the tribe Gossypieae, there should be more intensive sampling from *Cienfuegosia*, *Hampea*, and *Thespesia* sect. *Thespesia*. Furthermore, *Cephalohibiscus*

peekelii Ulbr. should also be included, as it is the sole member of the genus for which no data have been accumulated.

(ii) The use of faster evolving molecular data sets is necessary if we are to reveal phylogenetic relationships in groups of [*Hampea*, *Lebronnecia*, *Thespesia* sect. *Lampas*], [(*Hampea*, *Lebronnecia*, *Thespesia* sect. *Lampas*), *Thespesia* sect. *Thespesia*, and (*Gossypium*, *Kokia*, *Gossypioides*)]. At present, these polytomies remain unresolved.

(iii) This study suggested an unnatural taxonomic grouping of species in *Thespesia*, therefore, formal taxonomic revision of this genus is necessary.

(iv) For a better understanding of phylogenetic relationships among genomic groups in *Gossypium*, more data are needed.

(v) More taxonomic study is needed of species of Australian *Gossypium* sect. *Grandicalyx*, to address species delimitation and circumscription questions.

(vi) Because polymorphism exists in species of sect. *Grandicalyx*, it would be of interest to find out if this polymorphism is due to lineage sorting of ancient polymorphism, interspecies gene flow, or circumscription errors.

(vii) Since one species in sect. *Grandicalyx*, *Gossypium cunninghamii*, has an ancient hybridization in its evolutionary history, it would be of interest to address when hybridization occurred, which species were involved, whether or not there has been introgression of nuclear genes, and whether or not its isolated geographic distribution emerged as the result of the introgression.

(viii) Finally, further study is needed of ant-plant mutualism in Australian *Gossypium* sect. *Grandicalyx* to address relevant questions regarding seed dispersal, species distribution, patterns of genetic variation, the fate of seeds, diversity of ants, and specificity of ant species to any *Gossypium* species.

Many of the above questions could be answered by using more and faster evolving molecular data and/or more samples (e.g., i-v) whereas others need to encompass other types of studies (e.g., v-viii). An integrated

approach involving several disciplines is thus likely to be the most informative about this interesting tribe and genus.

LITERATURE CITED

ALEFELD, F. G. C. 1861. Ueber die Stellung der Gattung *Gossypium* und mehrer Anderer. Botanische Zeitung 19: 229-301.

BAIN, J. F., and R. K. JANSEN. 1996. Numerous chloroplast DNA polymorphisms are shared among different populations and species in the aureoid *Senecio* (Packera) complex. Canadian Journal of Botany 74: 1719-1728.

BAILEY, C. D., and J. J. DOYLE. 1997. The chloroplast *rpl2* intron and ORF184 as phylogenetic markers in the legume tribe Desmodieae. Systematic Botany 22: 133-138.

BALDWIN, B. G. 1993. Molecular phylogenetics of *Calycadenia* (Compositae) based on ITS sequences of nuclear ribosomal DNA: Chromosomal and morphological evolution reexamined. Americam Journal of Botany 80: 222-238.

BELLAROSA, R., V. DELRE, B. SCHIRONE, and F. MAGGINI. 1990. Ribosomal RNA genes in *Quercus* spp. (Fagaceae). Plant Systematics and Evolution 172: 127-139.

BENTHAM, G. and J. D. HOOKER. 1862-1867. *Genera plantarum.* vol. 1. Pp. 206-209. London: Reeve & Co.

BOBOLA, M. S., ECKERT, R. T., and KLEIN, A. S. 1992. Restriction fragment variation in the nuclear ribosomal DNA repeat unit within and between *Picea rubens* and *Picea mariana.* Canadian Journal of Botany 22: 255-263.

BOGLER, D. J., and B. B. SIMPSON. 1996. Phylogeny of Agavaceae based on ITS rDNA sequence variation. American Journal of botany 83: 1225-1235.

BOHS, L., and R. G. OLMSTEAD. 1997. Phylogenetic relationships in *Solanum* (Solanaceae) based on *ndhF* sequences. Systematic Botany 22: 5-17.

BRUBAKER, C. L., and J. F. WENDEL. 1994. Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). American Journal of Botany 81: 1309-1326.

CHASE, M. W., D. E. SOLTIS, R. G. OLMSTEAD, D. MORGAN, D. H. LES, B. D., MISHLER, M. R. DUVALL, R. A. PRICE, H. G. HILLIS, Y.-L. QUI, K. A. KRON, J.

H. Retting, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sytsma, H. J. Michaels, W. J. Kress, K. G. Karol, W. D. Clark, M. Hedren, B. S. Gaut, R. K. Jansen, K.-J. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, S. H. Strauss, Q.-Y. Xiang, G. M. Plunkett, P. S. Soltis, S. M. Swesen, S. E. Williams, P. A. Gardek, C. J. Quinn, L. E. Eguiarte, E. Golenberg, G. H. Learn, S. Graham, S. C. H. Barrett, S. Dayanandan, and V. A. Albert. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences rom plastid gene *rbcL*. Annals of Missouri Botanical Graden 80: 528-580.

Clark, L. G., W. Zhang, and J. F. Wendel. 1995. A phylogeny of the grass family (Poaceae) based on *ndhF* sequence data. Systematic Botany 20: 436-460.

Conant, D. S. 1995. The relationships of Papuasian Cyathaceae to New World tree ferns. American Fern Journal 85: 328-340.

Corner, E. J. H. 1976. The seeds of dicotyledons. 2 vols. New York: Cambridge University Press.

Crane, C. F., H. J. Price, D. M. Stelly, D. G. Czeshin, and T. D. MacKnight. 1993. Identification of the homeologous chromosome pair by *in situ* DNA hybridization to ribosomal RNA loci in meiotic chromosomes of cotton (*Gossypium hirsutum*). Genome 36: 1015-1022.

Cronn, R. C., X. Zhao, A. H. Paterson, and J. F. Wendel. 1996. Polymorphism and concerted evolutionin a tandomly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. Journal Molecular Evolution 42: 685-705.

Desfeux, C., S. Maurice, J.-M. Henry, B. Lejeune, and P.-H. Gouyon. 1996. Evolution of reproductive systems in the genus Silene. 263: 409-414.

Degener, O. 1934. *Kokia*. In *Flora Hawaiiensis; The new illustrated flora of the Hawaiian Islands*. Fam. 221., eds. O. Degener and L. D. Whitney. Honolulu.

DeJoode, D. R. and J. F. Wendel. 1992. Genetic diversity and origin of the Hawaiian Islands cotton, *Gossypium tomentosum*. American Journal of Botany 79: 1311-1319.

Downie, S. R., D. S. Katz-Downie, and K.-J. Cho. 1996. Phylogenetic analysis of Apiaceae subfamily Apioideae using nucleotide sequences

from the chloroplast *rpoC1* intron. Molecular Phylogenetics and Evolution 6: 1-18.

DOYLE, J. J., and A. H. D. BROWN. 1989. 5S Nuclear ribosomal gene variation in the *Glycine tomentella* polyploid complex (Leguminosae). Sytematic Botany 14: 398-407.

—, J. L. DOYLE, and J. D. PALMER. 1995. Multiple independent losses of two genes and one intron from legume chloroplast genomes. Systematic Botany 20: 22: 72-294.

—, —, J. A. BALLENGER, and J. D. PALMER. 1996. The distribution and phylogenetic significant of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae. Molecular Phylogenetics and Evolution 5: 429-438.

DUBCOVSKY, J., and J. DVORÁK. 1995. Ribosomal RNA multigene loci: Nomad of the Titiceae Genome. Genetics 140: 1367-1377.

DUMONT, A. 1887. Recherches sur l'anatomie comparée des Malvacées, Bombacacées, Tiliacées, Sterculiacées. Annales des Sciences Naturelles Seotiéne Serié 6: 129-246.

DUVALL, M. R., and B. R. MORTON. 1996. Molecular phylogenetics of Poaceae: An expanded analysis of *rbcL* sequence data. Molecular Phylogenetics and Evolution 5: 352-385.

EDLIN, H. L. 1935. A critical revision of certain taxonomic groups of the Malvales part II. New Phytologist 34: 122-143.

ENDRIZZI, J. E., E. L. TURCOTTE, and R. J. KOHEL. 1985. Genetics, cytology, and evolution of *Gossypium*. Advances in Genetics 23: 271-375.

FRYXELL, P. A. 1968. A redefinition of tribe Gossypieae. Botanical Gazette 129: 296-308.

—. 1969a. The genus *Hampea* (Malvaceae). Brittonia 21: 359-396.

—. 1969b. The genus *Cienfuegosia* Cav. (Malvaceae). Annals of the Missouri Botanical Garden 56: 179-250.

—. 1979. *The natural history of the cotton tribe*. College Station: Texas A & M University Press.

—. 1988. Malvaceae of Mexcico. Systematic Botany monograph vol. 25: 1-522.

—. 1992. A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea 2: 108-165.

—, L. A. Craven and J. McD. Stewart. 1992. A revision of *Gossypium* sect. *Grandicalyx* (Malvaceae), including the description of six new species. Systematic Botany 17: 91-114.

Fosberg, F. R. and M. H. Sachét. 1966. *Lebronnecia*, gen. nov. (Malvaceae) des Iles Marquises. Adansonia n.s. 6:507-510.

Gillies, A. C. M., and R. J. Abbott. 1996. Phylogenetic relationships in the genus *Stylosanthes* (Leguminosae) based on upon chloroplast DNA variation. Plant Systematics and Evolution 200: 193-211.

Gielly, L. and P. Taberlet. 1996. A phylogeny of the European gentians inferred from chloroplast *trnL* (UAA) intron sequences. Botanical Journal of the Linnean Society 120: 57-75.

Gottieb, L. D., and V. S. Ford. 1996. Phylogenetic relationships among the section *Clarkia* (Onagraceae) inferred from the nucleotide sequences of *PgiC*. Systematic Botany 21: 45-62.

Guo, J. H., D. Z. Skinner, and G. H. Liang. 1996. Phylogenetic relationships of *Sorghum* taxa inferred from mitochondrial DNA restriction fragment analysis. Genome 39: 1027-1034.

Harrison, R. E., J. J. Luby, and G. R. Furnier. 1997. Chlorolast DNA restriction fragment variation among strawberry (*Fragaria* spp.) taxa. Journal of American Society of horticultural Sciences 122: 63-68.

Holst-Jensen, A., L. M. Kohn, K. S. Jakobsen, and T. Schumacher. 1997. Molecular phylogeny and evolution of *Monilinia* (Sclerotiniaceae) based on coding and noncoding rDNA sequences. Americam Journal of Botany 84: 686-701.

Hoot, S. B. 1995a. Phylogenetic relations in *Anemone* (Ranunculaceae) based on DNA restriction site variation and morphology. Plant Systematics and Evolution [Supplement] 9: 295-300.

—. 1995b. Phylogeny of the Ranunculaceae baed on preliminary *atpB*, *rbcL*, and 18S nuclear ribosomal DNA sequence data. Plant Sytematics and Evolution [Supplement] 9: 241-251.

Hutchinson, J. B. 1947. Notes on the classification and distribution of genera related to *Gossypium*. New Phytologist 46: 123-141.

JOHNSON, L. A., J. L. SCHULTZ, D. E. SOLTIS, and P. S. SOLTIS. 1996. Monophyly and generic relationships of Polemoniaceae based on *matK* sequences. American Journal of Botany 83: 1207-1224.

KATAYAMA, H., and Y. OGIHARA. 1996. Phylogenetic affinities of the grasses to other monocots as revealed by molecular analysis of chloroplast DNA. Current Genetics 29: 572-581.

KEARNEY, T. H. 1951. The American genera of Malvaceae. American Midland Naturalist 46: 93-131.

KELCHNER, S. A., and L. G. CLARK. 1997. Molecular evolution and phylogenetic utility of the chloroplast rpl16 intron in Chusquea and the Bambusoideae (Poaceae). Molecular Phylogenetics and Evolution 8: 385-397.

KOHN, J. R., S. W. GRAHAM, B. MORTON, J. J DOYLE, and S. C. H. BARETT. 1996. Reconstruction of the evolution of reproductive characters in Pontederiaceae using phylogenetic evidence from chloroplast DNA restriction-site variation. Evolution 50: 1454-1469.

KNOX, E. R., and J. D. PALMER. 1995a. The origin of *Dendrosenecio* within the Senecioneae (Asteraceae) based chloroplast DNA evidence. American Journal of Botany 82: 1567-1573.

—, and —. 1995b. Chloroplast DNA variation and the recent radiation of the gigant senecio (Asteraceae) on the tall mountains of eastern Africa. Proceedings of National Academy of Sciences of the U.S.A. 92: 10349-10353.

KRON, K. A. 1996. Phylogenetic relationship of Empetraceae, Epacidaceae, Ericaceae, Monotropaceae, and Pyrolaceae: evidence from nuclear ribosomal 18S sequence data. Annals of Botany 77: 293-303

KUZOFF, R. K., J. A. SWEERE, D. E. SOLTIS, P. S. SOLTIS, and E. A. ZIMMER. 1998. The phylogenetic potential of entire 26S rDNA sequences in plants. Molecular Phylogenetics and Evolution 15: 251-263.

LA DUKE, J. C. and J. DOEBLEY. 1995. A chloroplast DNA based phylogeny of the Malvaceae. Systematic Botany 20: 259-271.

LEVY, F., J. ANTONOVIC, J. E. BOYNTON, and N. W. GILLHAM. 1996. A population genetic analysis of chloroplast DNA in *Phacelia*. Heredity 76: 143-155.

LIANG, H., and K. W. HILU. 1996. Application of the *MatK* gene sequences to grass systematics. Canadian Journal of Botany 74: 125-134.

LUO, H., and BOUTRY, M. 1995. Phylogenetic relationships within Hevea brasiliensis as deduced froma polymorphic nitochondrial DNA region. Theoretical and Applied Genetics 91: 876-884.

MASON-GAMER, R. J. and E. A. KELLOGG. 1996. Testing for phylogenetic conflict among molecular data sets in the Triticeae (Gramineae). Systematic Biology 45: 524-545.

—, K. E. HOLSINGER, and R. K. JANSEN. 1995. Chloroplast DNA haplotype variation within and among populations of *Coreopsis grandiflora* (Asteraceae). Molecular Biology and Evolution 12: 371-381.

MORI, N., N. T. MIYASHITA, T. TERACHI, and C. NAKAMURA. 1997. Variation in *coxII* intron in the wild ancestor species of wheat. Hereditas 126: 281-288.

MOUSADIK, A. E., and R. J. PETIT. 1996. Chloroplast DNA phylogeography of the argan tree of Morocco. Molecular Ecology 5: 547-555.

NEYLAND, R., and L. E. URBATSCH. 1996. Phylogeny of subfamily Epidendroideae (Orchidaceae) inferred from *ndhF* chloroplast gene sequences. American Journal of Botany 83: 1195-1206.

OHNISHI, O., and Y. MATSUOKA. 1996. Search for the wild ancestor of buckwheat II. Taxonomy of *Fagopyrum* (Polygonaceae) species based on morphology, isozymes, and cpDNA variablity. Genes and Genetic System 71: 383-390.

OLMSTEAD, R. G., and J. D. PALMER. 1994. Chloroplast DNA systematics: A review of methods and data analysis. American Journal of Botany 81: 1205-1224.

—, and —. 1997. Implications for the phylogeny, classification, and biogeography of *Solanum* from cpDNA restriction site variation. 22: 19-29.

—, and P. A. REEVES. 1993. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbcL* and *ndhF* sequences. Annals of Missouri Botanical Garden 82: 176-193.

PALMER, J. D. 1991. Plastid chromosome: structure and evolution. Cell Culture and Somatic Cell Genetics of Plants 7B: 5-53.

—. 1992. Mitochondrial DNA in plant systematics: Applications and limitation. In *Molecular systematics of plants*. P. S. Soltis, D. E.

Soltis, and J. J. Doyle (eds.). pp. 36-49. New York: Chapman and Hall.

Potter, D. S.,G. W. Andersen, and A. Robert. 1997. Phylogenetic relationships of the Raphidophyceae and Xanthophyceae as inferred from nucleotide sequences of the 18S ribosomal RNA gene. American Journal of Botany 84: 966-972.

Price, H. J., D. M. Stelly, T. D. McKnight, C. F. Scheuring, D. Raska, M. J. Michaelson, and D. Bergeey. 1990. Molecular cytogenetic mapping of a nucleolar organizer region in cotton. Journal of Heredity 81: 365-370.

Plunkett, G. M., D. E. Soltis, and P. S. Soltis. 1997. Clarification of the relationship between Apiaceae and Araliaceae based on *MatK* and *rbcL* sequence data. America Journal of Botany 84: 565-580.

Reddy, P., R. Appels, and B. R. Baum. 1990. Ribosomal DNA spacer-length variation in *Secale* spp. (Poaceae). Plant Systematics and Evolution 171: 205-220.

Reeves, R. G. 1936. Comparative anatomy of the seeds of cotton and other malvaceous plants II: Hibisceae. American Journal of Botany 23: 394-405.

Rodman, J. E., K. G. Karol, R. A. Price, and K. J. Sytsma. 1996. Molecules, morphology, and Dahlgren's expanded order Capparales. Systematic Botany 21: 289-307.

—, R. A. Price, K. Karol, E. Conti, K. J. Sytsma, and J. D. Palmer. 1993. Nucleotide sequences of the rbcL gene indicate monophyly of mustard oil plants. Annals of Missouri Botanical Garden 80: 686-699.

Sang, T., D. J. Crawford, and T. Stuessy. 1995. Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: Implication for biogeography and concerted evolution. Proceedings of the National Academy of Sciences USA 92: 6813-6817.

Savard, L., M. Michaud, and J. Bousquet. 1993. Genetic diversity and phylogenetic relationships between birches and alders using ITS, 18S

rDNA, and *rbcL* gene sequences. Molecular Phylogenetics and Evolution 2: 112-118.

SAVOLAINEN, V., R. SPICHIGER, and J.-F. MANEN. 1997. Polyphyletism of Celastrales deduced from a chloroplast noncoding DNA rgion. Molecular Phylogeny and Evolution 7: 145-157.

SEELANAN, T., A. SCHNABEL, and J. F. WENDEL. 1997. Congruence and consensus in the cotton tribe (Malvaceae). Systematic Botany 22: 259-290.

SEWELL, M. M., C. R. PARK, and M. W. CHASE. 1996. Intraspecific chloroplast DNA variation and biogeography of North American *Liriodendron* L. (Magnoliaceae). Evolution 50: 1147-1154.

SCHUMANN, K., AND M. GÜRKE. 1891-1892. Malvaceae I & II. In Flora Brasiliense, C. F. P. von Martius ed. pp. 12(3): 253-624.

SMALL, R. L., J. A. RYBURN, R. C. CRONN, T. SEELANAN, and J. F. WENDEL. 1998. The tortoise and the hare: Choosing between non-coding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. American Journal of Botany (in press).

SOLTIS, D. E., and R. K. KUZOFF. 1993. ITS sequence variation within and among populations of *Lomatium grayi* and *L. laevigatum* (Umbelliferae). Molecular Phylogenetics and Evolution 2: 166-170.

—, and P. S. SOLTIS. 1997. Phylogenetic relationships in Saxifragaceae sensu lato: a comparison of topologies based on 18S rDNA and rbcL sequences. American Journal of Botany 84: 504-522.

—, R. K. KUZOFF, E. CONTI, R. GORNALL, and K. FERGUSON. 1996. *MatK* and *rbcL* gene sequence data indicate that *Saxifraga* (Safragaceae) is polyphyletic. American Jornal of Botany 83: 371-382.

SPOONER, D. M., and R. T. CASTILLO. 1997. Reexamination of series relationships of South American wild potatoes (Solanaceae: *Solanum* sect. *Petota*): Evidence from chloroplast DNA restriction site variation. American Journal of Botany 84: 671-685.

STEPHENS, S. G. 1966. The potential for long range oceanic dispersal of cotton seeds. American Naturalist 100: 199-210.

STEWART, J. McD. 1995. Potential for crop improvement with exotic germplasm and genetic engineering. Challenging the Future:

Procedings of the World Cotton Research. Pp. 313-327. In: G.A. Constable & N.W. Forrester (eds.). Conference-1. CSIRO, Melbourne.

—, L. A. CRAVEN and J. F. WENDEL. 1997. A new Australian species of *Gossypium*. P. 448 in *Proceedings of Beltwide Cotton Conference*. National Cotton Council, Memphis, TN.

SUH, Y., L. B. THIEN, H. E. REEVE, and E. A. ZIMMER. 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. American Journal of Botany 80: 1042-1055.

SUN, Y., SKINNER, D. Z., LIANG, G. H., and S. H. HULBERT. 1994. Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. Theoretical and Applied Genetics 89: 26-32.

SYTSMA, K. J., and B. A. SCHAAL. 1985. Phylogenetics of the *Lisianthisus skinneri* (Gentianaceae) species complex in Panama utilizing DNA restriction fragment analysis. Evolution 39: 594-608.

TERRY, R. G., and G. K. BROWN. 1997. Phylogenetic relationships in subfamily Tillandsioideae (Bromeliaceae) using *ndhF* sequences. Systematic Botany 22: 333-345.

—, —, and R. G. OLMSTEAD. 1997. Reexamination of subfamilial phylogeny in Bromeliaceae using comparative sequencing of the plastid locus *ndhF*. American Journal of Botany 84: 664-670.

TSUMURA, Y., T. KAWAHARA, R. WICKNESWAI, and Y. YOSHIMURA. 1996. Molecular phylogeny of Dpterocarpaceae in Southeast Asia using RFPL of PCR-amplified chlorplast genes. Theoretical and Applied Genetics 93: 22-29.

WALLACE, R. S., and J. H. COTA. 1996. An intron loss in the chloroplast gene *rpoC1* supports a monophyletic origin for the subfamily Cactoideae of the Cactaceae. Current Genetics 29: 275-281.

WENDEL, J. F. 1989. New World tetraploid cottons contain Old World cytoplasm. Proceedings of the National Academy of Sciences USA 86: 4132-4136.

—, and V. A. ALBERT. 1992. Phylogenetics of the cotton genus (*Gossypium*): Character-state weighted parsimony analysis of chloroplast DNA

restriction site data and its systematic and biogeographic implications. Systematic Botany 17: 115-143.

—, C. L. BRUBAKER, and A. E. PERCIVAL. 1992. Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. American Journal of Botany 79: 1291-1310.

—, A. SCHNABEL, and T. SEELANAN. 1995a. An unusual ribosomal DNA sequence from Gossypium gossypioides reveals ancient, cryptic, intergenomic introgression. Molecular Phylogenetics and Evolution 4: 298-313.

—, A. SCHNABEL, and T. SEELANAN. 1995b. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proceedings of the National Academy of Sciences U.S.A. 92: 280-284.

—, J. McD. STEWART and J. H. RETTIG. 1991. Molecular evidence for homoploid reticulate evolution in Australian species of *Gossypium*. Evolution 45: 694-711.

WAGSTUFF, S. J., and R. G. OLMSTEAD. 1997. Phylogeny of Labiatae and Verbenaceae inferred from *rbcL* sequences. Systematic Botany 22: 165-179.

WOJCIECHOWSKI, M. F., M. J. SANDERSON, B. G. BALDWIN, and M. J. DONOGHUE. 1993. Monophyly of aneuploid *Astragalus* (Fabaceae): Evidence from nuclear ribosomal DNA internal transcribed spacer sequences. Americam Journal of Botany 80: 711-722.

WOLF, A. D., W. J. ELISENS, L. E. WATSON, and C. W. DEPAMPHILIS. 1997. Using restriction-site variation of PCR-amplified cpDNA genes for phylogenetic analysis if tribe Cheloneae (Schrophulariaceae). Americam Journal of Botany 84: 555-564.

YAKAWA, T., and K. UEHARA. 1996. Vegetative diversification and radiation in subtribe Dendrobiinae (Orchidaceae): evidence from chloroplast DNA phylogeny and anatomical characters. Plant Systematics and Evolution 201: 1-14.

YUAN, Y.-M., P. KÜPFER, and J. J. DOYLE. 1996. Infrageneric phylogeny of the genus *Gentiana* (Gentianaceae) inferred from nucleotide sequences of the internal transcribed spacers (ITS) of nuclear ribosomal DNA. Americam Journal of Botany 83: 641-652.

ZOMLEFER, W. B. 1994. Guide to flowering plant families. Chapel Hill: The University of North Carolina Press.

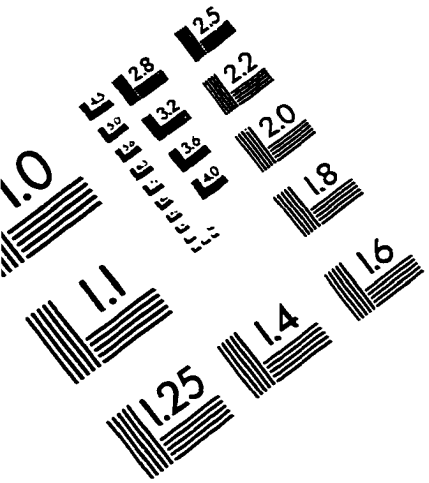
ACKNOWLEDGMENTS

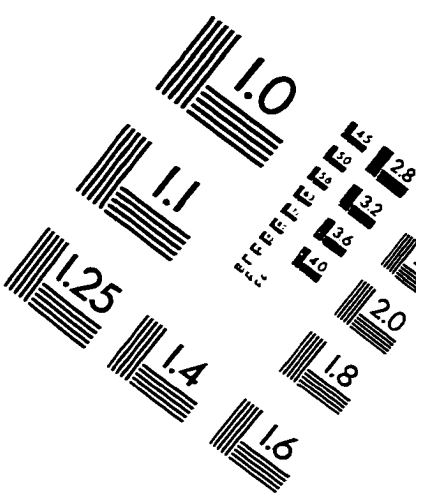
Most importantly, I would like to thank my co-major advisor, Dr. Jonathan F. Wendel who gave me an opportunity to work on this research. He has been a very good advisor, a good friend, and an excellent mentor. My success in the lab and my development as a scientist has been largely due to Jonathan. I also owe Curt many thanks for his suggestion to work on cotton (and Jonathan), and his collaboration with me on my second research. I would like to thank Lynn G. Clark for taking charge as my co-major advisor, and providing helpful guidance during my final stage toward my graduation. I am really grateful for Randy Small's helpfulness on his excellent and hard work on characterizing *Adh* genes. I also would like to thank many big cotton guys including Paul Fryxell (University of Texas), James Mac Stewart (University of Arkansas), and Lyn A. Craven (CSIRO, Australia) who have provided plant materials for my research. I am also thankful for my committee, Rob Wallace, Fred Janzen, Lois Tiffany, and James Colbert. Also, Rich Cronn and Andy Schnabel who have engaged in many discussions during the course of my research; many thanks to you.

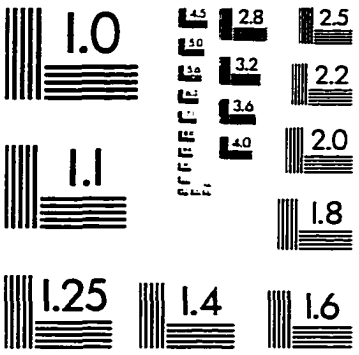
Life is not just in the lab, I enjoy socializing with you guys: Hugo Cota, Curt Brubaker, Andy Schnabel, Steve Dickie, Wendy Applequist, Phil Dykema, Donald Pratt, Randy Small, Julie Ryburn, Zhang Weiping. Also, thanks to Portia T.-W. Hsiau who has cared for me during my study.

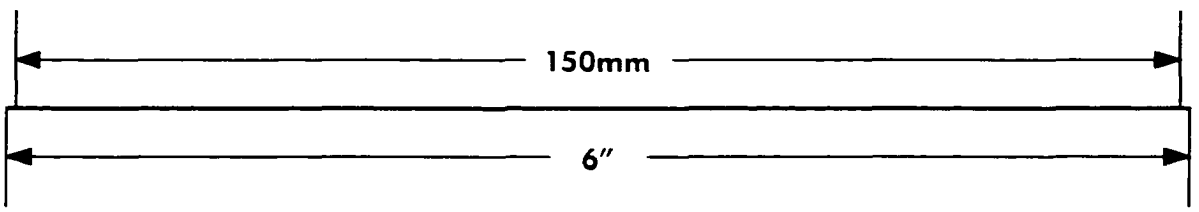
I would like to thank the Development and Promotion of Science and Technology Talents Project (DPST) of Thailand for generous scholarship support. I also would like to thank my family — Mom & Dad and my brother & sister —who are patiently supporting me emotionally.

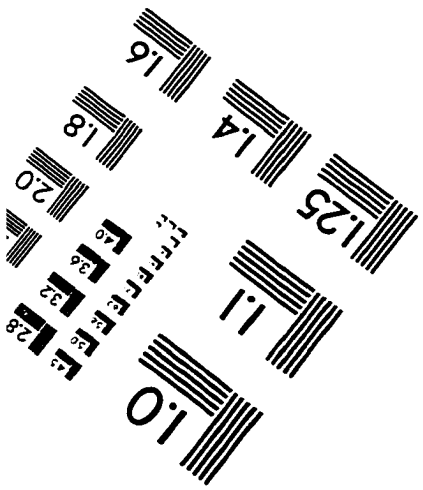
IMAGE EVALUATION TEST TARGET (QA-3)

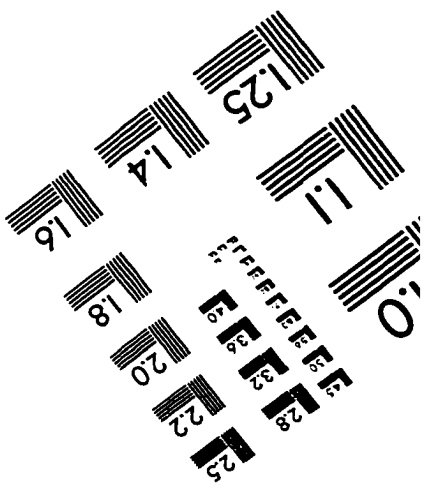












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989